Learning to lose

The role of input variability in the loss of V2

Marc Meisezahl

THE UNIVERSITY
of EDINBURGH

Doctor of Philosophy
The University of Edinburgh

2024

ABSTRACT

The loss of verb second (V2) in languages like English and French is often attributed to language-specific causes. One feature that is shared by all languages that have lost V2 is the increase in subjects in clause-initial position. This thesis explores how the variability in constituent types occurring in the clause-initial position affects the acquisition of V2. It is hypothesised that lower variability in the input affects the acquisition process negatively: Learners are unable to dissociate the clause-initial position from a specific constituent type, in particular subjects. The hypothesis is formulated against the backdrop of a domain-general fostering effect of variability on learning (cf. Raviv, Lupyan & Green 2022). The cognitive perspective that this approach takes allows the provision of a more uniform explanation for the changes that have occurred in individual languages. At the outset, variability in grammatical functions (e.g. subject, object) was stipulated to be the relevant domain. The hypothesis was then investigated with artificial language learning (ALL) experiments and a corpus study. Building on previous work, an ALL design was conceived that enables participants to sufficiently learn a V2 grammar in a short period of time (experiment 1 & 2). This design was put to use in three experiments with native speakers of English as participants. All three experiments compared languages with different distributions of clause-initial constituents: a high-variable language (i.e. with a uniform distribution of subjects, objects and adjuncts in clause-initial position), and two or three low-variable languages (i.e. with one constituent type dominating the clause-initial position). The experiments differed in the nature of the artificial language: a semi-artificial language combining English vocabulary with a V2 grammar (experiment 3), a fully artificial language (experiment 4) and a visual language that uses icons in lieu of lexical items (experiment 5) were used. The learning success in general was conditioned on the lexicon size such that a sufficiently sized lexicon was necessary for successful acquisition (experiment 4). Moreover, the results suggest that input variability indeed affects learning of a V2 grammar. Across experiments, a language where objects dominate the clause-initial position was the most difficult to learn. A language in which adjuncts dominate was learnt best in experiment 3, whereas no difference between such a language and a language with a uniform distribution of clause-initial elements was found for experiment 5. Although grammatical functions were initially defined as the relevant domain of variability, this could not be sustained in the light of the experimental findings. Instead, it is argued that learners can be sensitive to variability in grammatical functions and grammatical categories (e.g. NP, AdvP, CP). The uniform language exhibits high variability in terms of grammatical functions, while the adjunct-dominant language is characterised by high variability in grammatical categories. The advantage of the adjunct-dominant language in experiment 3 can be attributed to influence of participants' L1 (English).

This overall interpretation receives independent support from a large-scale corpus study showing a large proportion of adjuncts in initial position in German. Adjuncts entail high variability in grammatical functions and grammatical categories. The stability of the V2 property in German for over a millennium can be interpreted as direct consequence of a high proportion of clause-initial adjuncts. Taken together, the findings of this thesis provide support for the role the low variability in clause-initial constituent types played in the loss of V2. It further shows that a cognitive angle can be useful for finding uniform factors contributing to the loss of V2.

LAY SUMMARY

Among the languages of the world a host of different word orders can be found. A word order that is only very rarely attested is the so-called verb second (V2) word order. As the name suggests, the verb is placed in the second position of the clause. Only a single constituent — i.e. a single word or a group of words that form a syntactic unit (e.g. *the book*, *in the desert*) — can precede the verb. Although this resembles the word order of English, the crucial difference between English and V2 languages lies in the function of the clause-initial constituent. The preverbal position is confined to subjects in English, whereas only few restrictions apply to the same position in V2 languages. Examples of languages that are characterised by a V2 word order are German, Swedish and Icelandic. English also used to be a V2 language, but the V2 word order was replaced by the present-day subject-verb-object order in the 14th and 15th century. Other languages such as French and Portuguese share a similar fate. Interestingly, the loss of V2 in different languages proceeds in a similar fashion. On the one hand, word orders where the verb is preceded by more than one constituent can be found more frequently. On the other hand, the amount of sentences with subjects in initial position rises considerably. At the same time, V2 sentences in which a constituent other than the subject precedes the verb become increasingly rare. Research in linguistics and other domains of human cognition has shown that humans learn better (e.g. a motor skill) when they are exposed to variable stimuli during training. This raises the question whether the loss of variability in the position preceding the verb might have contributed to the loss of V2. This thesis investigates this question in a series of artificial language learning experiments. In this type of experiment, participants learn a miniature linguistic system that contains only the language features of interest. Such experiments thus offer the advantage that specific features can be studied in isolation without confounding effects from other aspects of the language. In the experiments presented in this thesis, English native speakers were taught V2 languages that differed only in the distribution of the different constituent types that precede the verb. Afterwards, it was tested whether participants apply the V2 rule to contexts that they have not seen before. The combined results of the experiments show that more variability in the constituents preceding the verb indeed benefits the acquisition of V2 languages. Lower variability, in turn, hinders the acquisition of V2 languages. The results are interpreted as evidence that the decrease in variability in the preverbal position can indeed contribute to the loss of V2. The findings of this thesis therefore suggest that properties of human cognition may have played a role in the loss of V2 across languages.

ACKNOWLEDGEMENTS

I am certainly not known for being outwardly emotional, but I feel immense gratitude for all the help I received and the experiences I could make while working on this thesis. First and foremost, I want to thank my incredible team of supervisors, Jennifer Culbertson and Simon Kirby. This thesis would not have been possible without their guidance, support and wisdom throughout my PhD journey. Despite the global challenges that have already cast a shadow over all of us in this decade, they have always remained supportive and their (virtual) doors open. I would have never been able to come to Edinburgh if it had not been for the funding I received from Jenny's ERC project *Syntax shaped by cognition* (grant agreement No 757643). Jenny's probing questions forcing me to continuously question my own assumptions were immensely helpful not only for my PhD project but also for my professional development. I am extremely grateful for her willingness to answer all my questions about the project and academia more generally and to read all my pieces of writing. Jenny also motivated me to finish writing my chapters when there was not yet light at the end of the tunnel. Simon's questions left a clear mark on my work by making me explore novel perspectives. Simon also reminded me of the bigger picture.

I also want to thank Mora Maldonado who has been on my supervision team in the first 18 months of my PhD. Mora helped to advance my experimental design and she was always open for questions.

I am grateful for the discussions I had with various people, while working on this thesis including Ailís Cournane, Heidi Getz, Caroline Heycock, Itamar Kastner, Dan Lassiter, Alexander Martin, Kenny Smith, Shira Tal, and George Walkden. Talking with these brilliant minds helped me to shape my thoughts and prompted me to explore aspects that I have not considered before. I also want to thank my two examiners, Caroline Heycock and Björn Lundquist, for the swift 'turn-around' and a very interesting, yet challenging discussion with interesting questions and comments.

I am grateful that I was given the opportunity to do a PhD at the CLE. I am still in awe of the great environment it provides. My PhD endeavour would not have been as much as fun as it was without my fellow CLE PhD students, past and present: Aislinn Keogh, Andres Karjus, Anna Kapron-King, Annie Holtz, Asha Sato, Elizabeth Pankratz, Fang Wang, Fausto Carcassi, Fiona Kirton, Henry Conklin, Juan Guerrero Montero, Jonas Nölle, Lauren Fletscher, Maisy Hallam, Patrick Kanampiu, Ponrawee Prasertsom, Shangmin 'Shawn' Guo, Svenja Wagner, Sydelle De Souza, Tamar Johnson, and Vlad Nedelcu. Our lunches, pub nights and retreats were a great source of joy. I also want to thank all of the other CLE members and visitors I had the pleasure of meeting in the last four years: Arianna Compostella, Danielle Naegeli, Marieke Schouwstra, Matt Spike, Rachel Kindellan, Richard Blythe, and Yevgen Matusevych. I also want to say

thank you to my Edinburgh office mates Alice, Bonnie, Carine, Richard, and Yajun for making going to the office much more fun once it was possible again. I also want so say a a big thank you to my new office mate Chiara Riegger for providing me what felt like an infinite chocolate source when I was in the last stages of writing this thesis.

Further, I want to thank Alisdair Tullo for his passionate assistance with python scripts and Sigurd Iversen for his help with creating the stimuli images of experiment 5. The fifth experiment in this thesis would not have been possible without my excellent models Cordula Englert, Roger Englert, Heidi Bartelt, Jeanne Englert, Jakob Roll, Donja Hirschmüller and Charlotte Englert who dedicated their well-deserved weekend to taking pictures with me in costumes that were almost unbearable to wear in the hot summer sun. Not a single grumbling in two days — Thank you all! I am grateful for the continuous support from the Postgraduate Office, especially Katie Keltie, Toni Noble, and Deanna Kaplan, and other administrative staff at the University of Edinburgh. I am also thankful for George Walkden's understanding of the challenges that come along with starting on a new project while simultaneously finishing a thesis.

My time in Edinburgh would not have been the same (inside and outside the pub) without my lovely pub quiz gang around Aislinn, Annie, Elizabeth, Henry, James, Maisy, Nik, and Sam. I am thankful for the fun evenings with Aisling, Alice, Bonnie, Chiara, Emma, Gerardo, Gwyn, Richard, and Sasha who reminded me that there is a life outside V2.

I want to say a massive thank you to Cordula and Roger who welcomed me into their home, not only once but twice during my PhD. They provided me with everything I needed, especially coconut chocolate and Paulaner Spezi. More importantly, however, they gave me something that I can call home again.

Last but not least, Charlotte. There are not enough words to say what I feel. You let me move abroad for my academic adventure and you helped me throughout this project. Your loving care, your scrupulous proof-reading and our animated discussions about all things V2 helped me to finish this thesis — it would not have been possible without you in my life. Danke für alles!

CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

1	First person
2	Second Person
3	Third person
ACC	Accusative
AUX	Auxiliary
CL	Clitic
DAT	Dative
DECL	Declarative
DEF	Definite
DIM	Diminutive
ERG	Ergative
EXPL	Expletive
FUT	Future
GEN	Genitive
IMPER	Impersonal verb form
INF	Infinitive
LOC	Locative
NEG	Negation
NF	Non-finite
NOM	Nominative
OBJ	Object
OV	Object voice
PAR	Partitive
PFV	Perfective
PL	Plural
PREP	Preposition
PRO	Proform
PRF	Perfect
PRS	Present
PRT	Particle
PST	Past
PTCP	Participle
Q	Question particle/marker
SG	Singular

ALL	Artificial language learning
CLD	Contrastive Left Dislocation
ClP	Classical Portuguese
DGS	German Sign Language
LES	Least Effort Strategy
ME	Middle English
MF	Middle French
OE	Old English
OF	Old French
OHG	Old High German
OP	Old Portuguese
PDG	Present Day German
PLD	Primary linguistic data
UG	Universal Grammar
V1	Verb first
V2	Verb second
V3	Verb third
V4	Verb fourth
VF	Verb final
WEIRD	Western, educated, industrialised, rich, democratic

CHAPTER 1

V2, ITS LOSS AND THE ROLE OF VARIABILITY

1.1 Introduction

Romansh, a Rhaeto-Romance language spoken in Switzerland (Canton of the Grisons), exhibits an intriguing word order. Different permutations of the same constituents are licensed by the grammar. SVO sentences, as in (1a), along with object-initial sentences (1b) and adjunct-initial sentences (1c) are permissible. Even sentences with clause-initial participles, exemplified in (1d), constitute grammatical utterances (Spescha 1989, Kaiser 2002, Liver 2010, Maurer-Cecchini 2021).[1,2]

(1) a. *La dunna* **ha** legiu il cudisch cun plascher. ROMANSH
the woman has read the book with pleasure
'The woman reads the book with pleasure.'
b. Il cudisch **ha** *la dunna* legiu cun plascher.
c. Cun plascher **ha** *la dunna* legiu il cudisch.
d. Legiu **ha** *la dunna* il cudisch cun plascher.
(Kaiser 2002: 2)

One striking feature that unifies all sentences in (1) is the position of the finite verb *ha*. Notice that the finite verb consistently occupies the second position of the clauses; that is, only one constituent precedes the verb. This peculiar word order is commonly referred to as verb second (V2) in the literature (Holmberg 2015). As the name suggests, the verb obligatorily occupies the second position in a clause. At the same time, the nature of the clause-initial constituent is virtually unconstrained. The apparent sensitivity to linear positions is cross-linguistically unattested: grammatical functions, not linear positions are often considered the determining factors in word order. When considered from a typological perspective, V2 appears to be a rare phenomenon (Holmberg 2015: 343). Only a small number among the currently spoken languages feature a V2 grammar; apart from Romansh, V2 is attested in almost all Germanic languages with the exception of English and a few other languages such as Kashmiri (Bhatt 1999), Berta (Andersen

[1] Throughout this thesis, different grammatical functions in sentences are visually encoded. Subjects are italicised, objects are underlined and finite verbs are represented in bold font.
[2] All Romansh examples are taken from the Sursilvan variety.

2017) and Tohono O'odham (Fitzgerald 2020). Even when historic stages of current non-V2 languages are included, only English, Welsh and most of the Romance language can be added to the list of V2 languages.

Interestingly, the languages that have lost V2 and those still exhibiting a V2 grammar shared many similarities previously. Apart from the frequent V2 sentences, non-V2 structures are attested in earlier stages as well. For instance, verb first (V1) sentences can be found both in Old English (OE) and Old High German (OHG). This can be seen in the examples in (2a) and (2b), respectively. At the same time, more than one constituent can also precede the finite verb, as exemplified in the examples in OE (3a) and OHG (3b).

(2) a. **Wæs** *Hæsten* þa þær cumen mid his herge OE
was Hæsten then there come with his host
'Hæsten then had come there with his host.'

b. **Was** *liutu filu* in flize, in managemo agaleize OHG
were people many in diligence in great effort
'There were many people in diligence, in great effort.'
(Hinterhölzl & van Kemenade 2012: 814)

(3) a. Hiora umtrymnesse *he* **sceal** ðrowian on his heotan OE
their weakness he shall atone in his heart
'He shall atone their weakness in his heart.'
(Haeberli 2002a: 90)

b. forlaaz senu dhir **uuer dant** *dhino suntea* OHG
forgiven you.DAT become your sins
'Your sins are forgiven.'
(Axel 2009a: 135)

Why does one language lose its V2 property, whereas the other has retained and even strengthened it (even more so when both are closely related)? Different explanations have been adduced for the loss of V2 — ranging from changes to inflection morphology (Haeberli 2002a), phonological (prosodic) changes (Adams 1987b, Galves & Paixão de Sousa 2017) to language contact (Kroch, Taylor & Ringe 2000). Even though the causes are generally language-specific, similarities can be observed as well. On the one hand, all languages that have lost V2 showed an increase in subject-initial clauses (van Kemenade & Westergaard 2012, Steiner 2014, Meelen 2016). On the other hand, many theories allude to some form of learning to explain the demise of V2 (Roberts 1993, Willis 1998, Steiner 2014). The goal of this thesis is to spell out what role learning plays and how learning interacts with the distribution of different clause-initial elements. This is inspired by learning models that connect language change to language acquisition (Yang 2002, Lightfoot 1999, Westergaard 2009b). Based on experimental work showing that variability in the input fosters learning, I hypothesise that the loss of V2 is contingent on the variability in the clause-initial position. If the variability is too low, V2 will be lost. I specifically propose that low variability prevents learners from forming representations where no fixed association between the clause-initial position

and specific grammatical categories (e.g. NP, PP) or grammatical functions (e.g. subject, object) exist.

The purpose of the present chapter is to lay the foundations for this thesis. First, the phenomenon under investigation will be introduced, including a brief summary of theoretical analyses (§1.2). Next, the loss of V2 in different languages will be scrutinised more closely (§1.3), before three models of learning will be sketched (§1.4). In §1.5, the role of variability will be examined and the hypothesis under investigation in this thesis will be derived. Finally, the artificial language learning method will be introduced in §1.6 before the chapter concludes with a brief summary and an overview of this thesis (§1.7).

1.2 V2: An overview

I will first provide a descriptive definition of V2 (§1.2.1) before listing V2 languages in the past and present (§1.2.2). The section will conclude with a brief overview of three different types of theoretical analyses of the studied phenomenon (§1.2.3).

1.2.1 Towards a descriptive definition of V2

The example in (1), repeated here in (4) for convenience, showcases again the V2 word order pattern in Romansh. The finite verb forms the second constituent of the clause irrespective of the nature of the clause-initial constituent. That is, the verb always comes second when the subject (4a), the object (4b) and even non-arguments (4c)–(4d) are realised clause-initially.

(4) a. *La dunna* **ha** legiu il cudisch cun plascher. ROMANSH
the woman has read the book with pleasure
'The woman reads the book with pleasure.'
b. Il cudisch **ha** *la dunna* legiu cun plascher.
c. Cun plascher **ha** *la dunna* legiu il cudisch.
d. Legiu **ha** *la dunna* il cudisch cun plascher.
(Kaiser 2002: 2)

Although I will shortly discuss some exceptions (cf. also (2)–(3)), the following broad descriptive definition of V2 can be provided: The verb is obligatorily realised in the second position of the clause and can only be preceded by one constituent whereby few or no restrictions apply as to the grammatical function or category of the clause-initial element (Vikner 1995: 39, Kaiser 2002: 1, Holmberg 2015: 342, 347, Freitag 2021: 7). The obligatory placement of the verb in the second position becomes evident when (4) and (5) are compared. In (5a)–(5c) two constituents occupy the clause-initial position rendering the sentences ungrammatical. That is, if constituents other than the subject precede the verb, the subject and the verb need to invert (cf. (4b)–(4d)).

(5) a. **La dunna* il cudisch **ha** legiu cun plascher. ROMANSH
b. *Cun plascher *la dunna* **ha** legiu il cudisch.
c. *Il cudisch cun plascher **ha** *la dunna* legiu.
(Kaiser 2002: 2)

Although all previous examples featured exclusively declarative sentences, V2 is not necessarily confined to declaratives. In Romansh, V2 is also required in *wh*-interrogatives (Spescha 1989: 594, Liver 2010: 147, Hack & Kaiser 2013) and may occur in certain types of subordinate clauses (Grünert 2018, Meisezahl 2019). This is illustrated in (6) as well as (7) and (8), respectively.

(6) Tgei **manegia** *tiu patrun*? ROMANSH
what means your boss
'What does your boss think?'
(Hack & Kaiser 2013: 146)

(7) Cura **eis** *ella* morta? ROMANSH
when is she died
'When did she die?'
(Liver 2010: 147)

(8) […] che lu **seigi** *ei* in bienton pli ruasseivel ROMANSH
[…] that then be it a piece more quietly
'[…] that it would be somewhat quieter then'
(Grünert 2018: 28)

Freitag (2021: 9) notes that the presence or absence of a V2 order is contingent on the clause-type and is language-dependent. This can be seen when Romansh and Yiddish — another V2 language — are compared. While polar questions in Romansh are V1 structures (9) (Spescha 1989: 594, Liver 2010: 147), the word order can alternate between V1 (10a) and V2 (10b) in Yiddish polar questions (Diesing 1990: 55–56).

(9) **Eis** *el* aunc cheu? ROMANSH
is he still here
'Is he still here?'
(Liver 2010: 147)

(10) a. **Hot** *er* gezen Maxn? YIDDISH
has he seen Max
'Has he seen Max?'
b. Tsi **hot** *er* gezen Maxn?
Q has he seen Max
(Diesing 1990: 55–56)

The same applies to permissible constituent types in clause-initial position. Different V2 languages license different clause-initial constituents (Freitag 2021). This can be illustrated with the contrast between two further V2 languages, Swedish and Icelandic. In Swedish, eventive VPs can be realised in the clause-initial position if the verb is

not stative (Källgren & Prince 1989: 48–49): The well-formedness of (11a) with a clause-initial VP of an eventive verb stands in stark contrast to the one (11b) where the clause-initial VP is headed by a stative verb. Icelandic on the other hand rules out any fronted verb, both as VP (12a) and as sole verb (12b) (Thráinsson 2007: 344, 349). Despite those differences, it is worth pointing out that all V2 languages permit subjects, objects, adverbials and *wh*-elements as clause-initial constituents (Holmberg 2015).

(11) a. Läser boken **gör** *han* nu. SWEDISH
reads book.DEF does he now
'Reading the book he is now.'
(Källgren & Prince 1989: 47)

b. *Kan svenska **gör** *Kari*.
knows Swedish does Kari
'Knowing Swedish, Kari does.'
(Källgren & Prince 1989: 49)

(12) a. *Lesa allar bækurnar **mun** *hún* ICELANDIC
read.INF all books.DEF will she
'She will read all the books'
(Thráinsson 2007: 349)

b. *Lesa **munu** *strákarnir* bækunar.
read.INF will boys.DEF books.DEF
'The boys will read the books.'
(Thráinsson 2007: 344)

Even though V2 is normally strictly followed in the context it occurs, context-specific exceptions have been noted for all languages. Coming back to Swedish, the V2 rule is optional with certain focus adverbs such as *bara* 'only' and *till och med* 'even', but also with *kanske* 'maybe' (13) (Holmberg 2015: 355). Furthermore, the left dislocation of constituents results in linear verb third (V3) orders (Holmberg 2015: 354). This can be seen in the Contrastive Left Dislocation (CLD) in (14) where the dislocated form is resumed by a proform in the preverbal position.

(13) Kanske (**kommer**) *han* inte (**kommer**). SWEDISH
maybe comes he no comes
'Maybe he's not coming.'
(Holmberg 2015: 355)

(14) För två veckor sen, då **köpte** *Johan* sin första bil. SWEDISH
for two weeks ago then bought Johan his first car
'Two weeks ago Johan bought his first car.'
(Holmberg 2015: 354)

This section focused on a small set of languages. However, Romansh, Yiddish, Swedish and Icelandic are not the only V2 languages. The next section will therefore provide an overview of the languages for which a V2 grammar has been proposed.

1.2.2 Languages with V2 order

When considered from a typological perspective, V2 constitutes a rare word order phenomenon (Holmberg 2015: 343). That is, only a relatively small number of languages have been described as V2 language. Traditionally, V2 has been mostly associated with the Germanic languages (den Besten 1989, Vikner 1995).[3] All Germanic languages — with the notable exception of English which is a SVO language (e.g. Vikner 1995, Yang 2002) — exhibit a V2 grammar, as the following examples demonstrate for German (15a), Norwegian (15b) and Faroese (15c). Although the SVO order of English may superficially resemble V2 sentences (16a), preposing of non-subjects to the clause-initial position renders sentences ungrammatical (16b) or the subject and verb fail to invert yielding V3 orders (16c).

(15) a. Mit Kartoffelsalat **schmecken** *Maultaschen* am besten. GERMAN
with potato salad taste maultaschen at.the best
'Maultaschen taste best when served with potato salad.'

b. I fjor **skrev** *hun* to bøker. NORWEGIAN
last year wrote she two books
'Last year she wrote two books.'
(Westergaard 2021a: 359)

c. Í ovurmorgin **hevur** *Karin* føðingardag. FAROESE
on day.after.tomorrow has Karin birthday
'The day after tomorrow, Karin has her birthday.'
(Heycock, Sorace & Hansen 2010: 65)

(16) a. *The bold mouse* **pursues** the hikers in Svensby.
b. *The hikers **pursues** *the bold mouse* in Svensby.
c. In Svensby *the bold mouse* **pursues** the hikers.

Among the Romance languages, the Dolomitic Ladin varieties Badiot (17) and Gherdëina (18) are the only varieties besides Romansh that feature a V2 word order (Poletto 2000, 2002, Kaiser 2002–2003, Kaiser & Hack 2013, Casalicchio & Cognola 2018, 2020). Further V2 languages within the Indo-European language family are Breton (Celtic; Schafer 1995, Borsley & Kathol 2000, Jouitteau 2007, Kennard 2018), Kashmiri (Indo-Aryan; Bhatt 1999, Manetta 2011, 2021) and the Kotgarhi and Koci dialects of Himachali (Indo-Aryan; Hendriksen 1990). The V2 property for each of these languages is illustrated in (19)–(21).

(17) Te botëga **à** tres *la* *mama* cumprè la farina. BADIOT
in shop has always the mum bought the flour
'Mum has always bought the flour in the shop.'
(Casalicchio & Cognola 2020: 603)

[3] An interesting word order pattern has been noted for Wymysorys, a German(ic) minority variety spoken in Wilamowice, a town in southern Poland. According to Andrason (2020), two word order systems co-exist in the grammar. On the one hand, speakers can access a Germanic V2 word order. On the other hand, speakers may choose a Slavonic system that allows a relatively free word order. The latter system is the result of close contact with Polish (Andrason 2020).

(18) La lëtra tla zaita l'**à** scrita *la l'oma*. GHERDËINA
the letter in.the newspaper her.OBJ.CL = has written the mum
'It is the mum who wrote the letter to the newspaper.'
(Casalicchio & Cognola 2018: 84)

(19) Al levr a **lennar** *Anna*. BRETON
the book PRT read Anna
'It was the book that Anna read.'
(Borsley & Kathol 2000: 667)

(20) Darvaaz **mustroov** *rameshan*. KASHMIRI
door opened Ramesh.ERG
'It was the door that Ramesh opend.'
(Bhatt 1999: 85)

(21) a. teu **kɔru** dziua laɪo KOTGARHI
that I-do my-mind fixing
'That (work) I do dilligently.'
b. tabːe **suntse** daljia *duǰːe tərkɪːb* KOCI
then was-devised by-the-miser another plan
'Then, another plan was devised by the miser'
(Hendriksen 1990: 162)

Outside the Indo-European language family, V2 is even less frequently attested. Estonian (Finno-Urgic), exemplified in (22), has been often adduced as example for a V2 language (Ehala 2006, Sahkai & Tamm 2019, Holmberg, Sahkai & Tamm 2020), although recent work suggests that this only applies to written Estonian (Vihman & Walkden 2021).[4] Another language which displays V2 patterns — at least in some contexts — is the Tupi language Karitiana (23) (Storto 1999, 2003, 2020). In addition to (written) Estonian and Karitiana, the Uto-Aztecan language Tohono O'odham (Miyashita 2006, Fitzgerald 2020) and the Nilo-Saharan languages Berta (Andersen 2017) and Dinka (van Urk 2015) have been described as V2 languages. Examples are provided in (24)–(26).

(22) Pühapäeviti **küpsetab** *Mari* tavaliselt kooki. ESTONIAN
on.Sundays bake.3SG Mari usually cake.PAR
'Mary usually bakes cakes on Sundays.'
(Holmberg, Sahkai & Tamm 2020: 439)

(23) Dibm **y-taka-tat-i** *yn* KARITIANA
tomorrow 1-DECL-go-FUT I
'I will go tomorrow.'
(Storto 2020: 325)

(24) Cúhug 'óidam **'at-t** cíkp. TOHONO O'ODHAM
last.night during 1PL.AUX-PFV to.work.PFV
'We worked all night long.'
(Fitzgerald 2020: 754)

[4] Vihman & Walkden (2021) propose a V3 analysis for spoken Estonian.

(25) ʔìŋíʃ **gìi** *níŋè* dùk'à θáŋ. BERTA
pot make.PRS woman porridge PREP.PRO
'The woman is cooking porridge with the pot.'
(Andersen 2017: 105)

(26) Cuu̱in **à-cu̱i** *Áyèn* câam nɛ̀ păal. DINKA
food 3SG-PRF.OV Ayen.GEN eat.NF PREP knife
'Food, Ayen has eaten with a knife.'
(van Urk 2015: 61)

When shifting the focus from a synchronic to a diachronic perspective, the status of V2 as a typologically rare phenomenon persists. However, the list of V2 languages can be extended by a small set of languages that exhibited a V2 grammar at some point in their history, but lost it at a later stage. One of the languages that falls into this group is English. Before losing V2 during the Middle English period, English was on a par with the other Germanic languages in terms of the main clause word order (e.g. van Kemenade 1987, Roberts 1996, Fischer et al. 2001, Westergaard 2009d, Haeberli & Ihsane 2016).[5] A similar development has been noted for Welsh in that affirmative main clauses exhibited a V2 order during the Middle Welsh period (Willis 1998, 2007, Meelen 2016, 2020). The last group of languages for which a V2 grammar has been proposed are the medieval Romance languages (Benincà 1995, 2006, Wolfe 2018c). This includes French (e.g. Adams 1987a,b, Roberts 1993, de Andrade 2018, Labelle & Hirschbühler 2018, Klævik-Pettersen 2019, Wolfe 2021), Occitan (Wolfe 2018a), Italian (Poletto 2014), Venetian (O. Singh 2021), Sicilian (Wolfe 2016, 2020), Spanish (Fontana 1993, 1997, Wolfe 2015b) and Portuguese (Ribeiro 1995, Galves 2018, 2020, de Andrade 2018). Possibly the only diachronic exceptions in terms of V2 among the Romance language are Sardinian (Wolfe 2015a, 2016, 2018c) and Catalan (Pujol i Campeny 2018).[6] Old Sardinian at least has retained a word order already present in Late Latin which in itself shows signs of an emerging V2 grammar (Wolfe 2015a, 2016, 2018c, Ledgeway 2017). It should also briefly be acknowledged that the analysis of Medieval Romance as V2 has not unanimously been accepted. The main argument for proponents of a non-V2 analysis lies in the attestation of a significant number of non-V2 constructions, contrasting with strict V2 languages such as German (Kaiser 2002, Kaiser & Zimmermann 2011, Rinke 2009, Sitaridou 2011, 2012, Martins 2019).[7] Scholars who consider the Medieval Romance languages to be V2 languages sometimes refer to them

[5] The V2 word order of English and some other previous V2 languages is discussed in more detail in §1.3.

[6] The fact that all of the Romance V2 languages except for Romansh and Dolomitic Ladin have lost their V2 grammar raises intriguing questions with respect to the status of the V2 grammars in those Rhaeto-Romance varieties. That is, do these V2 grammars constitute a remnant of an earlier cross-Romance system or should they rather be interpreted as independent developments? Kaiser & Hack (2013: 86) speculate that the close contact with German reinforced existing tendencies for a V2 order. However, due to the sparsity of historical data for Rhaeto-Romance, providing a definitive answer for those questions might prove a challenge.

[7] See Wolfe (2018c) for a discussion.

as *relaxed* V2 languages (Cognola 2015, 2019).[8] As the next section will show however, seemingly different types of V2 languages could be unified under certain theoretical syntactic analyses.

1.2.3 Generative approaches to V2

Hitherto in this thesis, V2 has been treated solely on a descriptive level. Within the theoretical syntactic literature however, V2 has played a significant role in the development of syntactic theories. Although the goal of this thesis is *not* to develop a theoretical account of V2, theoretical accounts have underpinned diachronic accounts of V2. An understanding of these theories is therefore necessary. This section will provide a brief and non-exhaustive overview of different syntactic accounts of V2. Specifically, symmetrical analyses (§1.2.3.1), asymmetrical analyses (§1.2.3.2) and analyses following Rizzi (1997) in invoking an articulated left periphery (§1.2.3.3) are discussed.[9] All of the analyses outlined here share the fundamental notion of V2 as a derived phenomenon. The main difference between the different strands lies in the targeted positions of individual constituents.

1.2.3.1 Symmetrical analyses

Symmetrical analyses, as proposed for instance by Vikner (1995) and Holmberg & Platzack (1995), share the fundamental insight that V2 is a derived phenomenon. More precisely, V2 is decomposed into two distinct movement operations: head-movement of the finite verb to C and phrasal movement of the initial constituent to the specifier of CP. Crucially, the fronting of the sentence-initial constituent must proceed via Ā-movement as SpecCP can host both arguments and non-arguments (as extensively illustrated by previous examples).[10,11] The fact that all clause-initial constituents are argued to occupy the same syntactic position has motivated the name for this type of analyses: subjects and and non-subjects stand in symmetrical relation to each other. This account offers a straightforward explanation for the position of the finite verb as well as the constraint on the number of constituents preceding the verb: The clause-initial constituent and the finite verb target the highest projection in the structure where only one landing position, namely SpecCP, is available for XPs. The second position of the verb in the linearisation on the other hand follows from the head-initial nature of CP. The movement of the verb

[8] Some of the modern V2 languages can also be classified as relaxed V2 languages. For instance, the Dolomitic Ladin varieties as well as Mòcheno and Cimbrian (two German(ic) minority varieties in Northern Italy) have been analysed as relaxed V2 languages (Cognola 2013, 2015, 2019).

[9] Other proposals for the analysis of V2 include e.g. remnant movement (Müller 2004). See Holmberg (2015) for an overview.

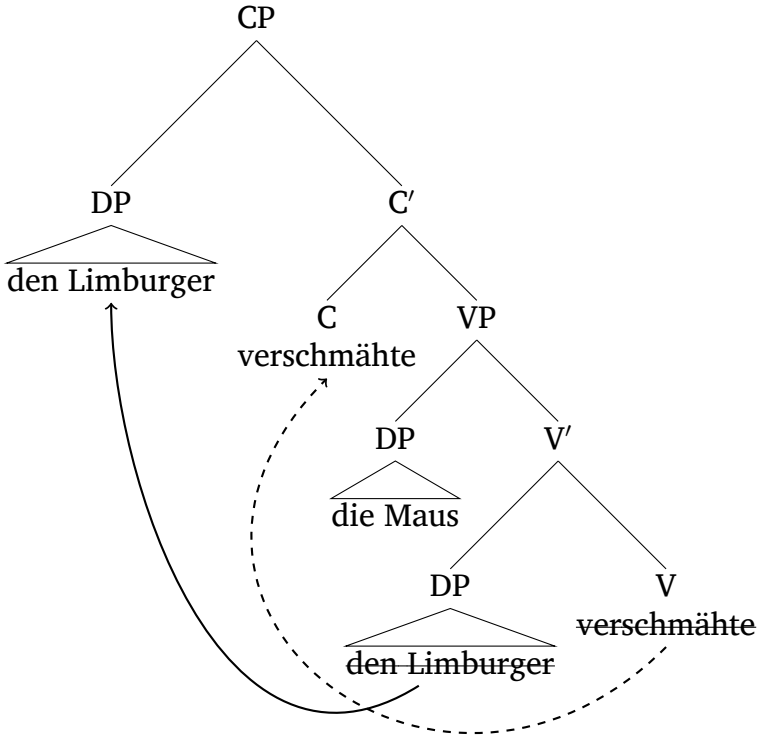
[10] An important consequence of this analysis is that V2 languages have underlying word orders. As will become clear shortly, Dutch and German are SOV languages for example.

[11] A related analysis assumes that the clause-initial constituent and the finite verb target SpecIP and I respectively (Diesing 1990, Pintzuk 1991, Santorini 1992). That is, SpecIP can be the target of both A-movement and Ā-movement. This analysis was prompted by the observation for Yiddish and Icelandic that V2 generally occurs in main *and* embedded contexts.

and the clause-initial constituent are illustrated in the simplified tree in (27), following Freitag (2021: 52).

(27) a. Den Limburger **verschmähte** *die Maus*. GERMAN
the.ACC limburger cheese spurned the.NOM mouse
'The mouse spurned the limburger cheese.'

b.



Symmetrical analyses build on seminal work by den Besten (1989) who observed that movement of the finite verb and lexical complementisers stand in complementary distribution in Dutch and German.[12] In other words, verbs are only realised in second position if no lexical complementiser is present. This can readily been seen in the example in (28). The auxiliary *hat* 'have' in (28a) is realised in the clause-final position when the complementiser *dass* 'that' occupies the clause-initial position. However, when the complementiser is omitted, as in (28b), the auxiliary occurs in the second position. Crucially, an overt complementiser in combination with verb movement results in an ungrammatical utterance as (28c) exemplifies. A straightforward conclusion that can thus be drawn from this data is that the verb and the complementiser must reside in the same syntactic position.

[12]The proposal of den Besten (1989) has been in circulation since 1977 and was first published in 1983 with two addenda compared to the original version. The version referenced here was published as part of den Besten's PhD dissertation with four further notes.

(28) The cheese monger says …

a. …dass *die Maus* den Käse angeknabbert **hat** GERMAN
…that the.NOM mouse the.ACC cheese nibbled has
'…that the mouse nibbled at the cheese.'

b. …*die Maus* **hat** den Käse angeknabbert.

c. *…dass *die Maus* **hat** den Käse angeknabbert.

The complementary distribution of finite verbs and lexical complementisers is not the only evidence for symmetrical analyses though. The placement of pronouns can serve as further evidence for the positional identity of complementisers and finite verbs in V2 languages (den Besten 1989: 25–34). Dutch features two sets of subjects pronouns — a set of strong pronouns and a set of weak pronouns. In V2 clauses, weak pronouns must be placed right-adjacent to the finite verb as the sentence becomes ungrammatical otherwise (29a).[13] That this is only a restriction on weak pronouns can be seen by the absence of comparable distributional restrictions for the strong equivalents in (29b). Relevant for the current argument is the syntactic behaviour of weak pronouns relative to lexical complementisers. The example in (30a) demonstrates that weak pronouns must be right-adjacent to the complementiser. Again, this restriction does not apply to the strong equivalents (30b). That is, the placement of the weak pronouns provides further support for the hypothesis that finite verbs and complementisers compete for the same position.

(29) a. Waarom **was** *(*ze*) gisteren (**ze*) ziek? DUTCH
why was she yesterday she ill
'Why was she ill yesterday?'

b. Waarom **was** (*zij*) gisteren (*zij*) ziek?
why was she yesterday she ill
'Why was she ill yesterday?'
(den Besten 1989: 26)

(30) a. …dat *(*ze*) gisteren (**ze*) ziek **was** DUTCH
…that she yesterday she ill was
'…that she was ill yesterday.'
(den Besten 1989: 25)

b. …dat (*zij*) gisteren (*zij*) ziek **was**
…that she yesterday she ill was
'…that she was ill yesterday.'
(den Besten 1989: 26)

To motivate the movement of the finite verb to C in the absence of an overt complementiser, den Besten (1989: 90–93) proposes that C hosts a feature that is also borne by the finite verb. In the absence of a complementiser, the feature in C triggers

[13] This constraint on subject pronoun placement does not only hold for Dutch but can be found in other West Germanic languages as well with the exception of West Flemish (Haeberli 2002a: 95).

verbal movement. In the proposal of den Besten (1989), this feature is [TENSE].[14] The movement of a constituent to SpecCP in turn is the result of a *constituent preposing rule* (den Besten 1989: 30–31). In more recent proposals, this rule is replaced by a movement-inducing feature. Roberts (2004) proposes that C (Fin in their analysis) bears an [EPP]-feature which requires the specifier of a phrase to be filled, thereby triggering movement of an XP to the clause-initial position.[15] The next strand of analyses — i.e. asymmetrical analyses — shares many features with symmetrical analyses but diverges with respect to the position of subjects.

1.2.3.2 Asymmetrical analyses

The asymmetrical analyses developed by Travis (1984, 1991) and Zwart (1997) deviate from symmetrical analyses in that different structural configurations are assumed for subject-initial and non-subject-initial clauses. Non-subject-initial sentences are analysed as similar to symmetrical analyses sketched in §1.2.3.1: The verb undergoes V-to-C movement and the clause-initial constituent is moved to SpecCP. The crucial difference between the two types of analyses lies in the analysis of subject-initial sentences. Unlike non-subject-initial sentences, the verb and the subject target a position in the IP-domain. The two configurations are illustrated in (31) and (32) for an object-initial and subject-initial sentence, respectively.

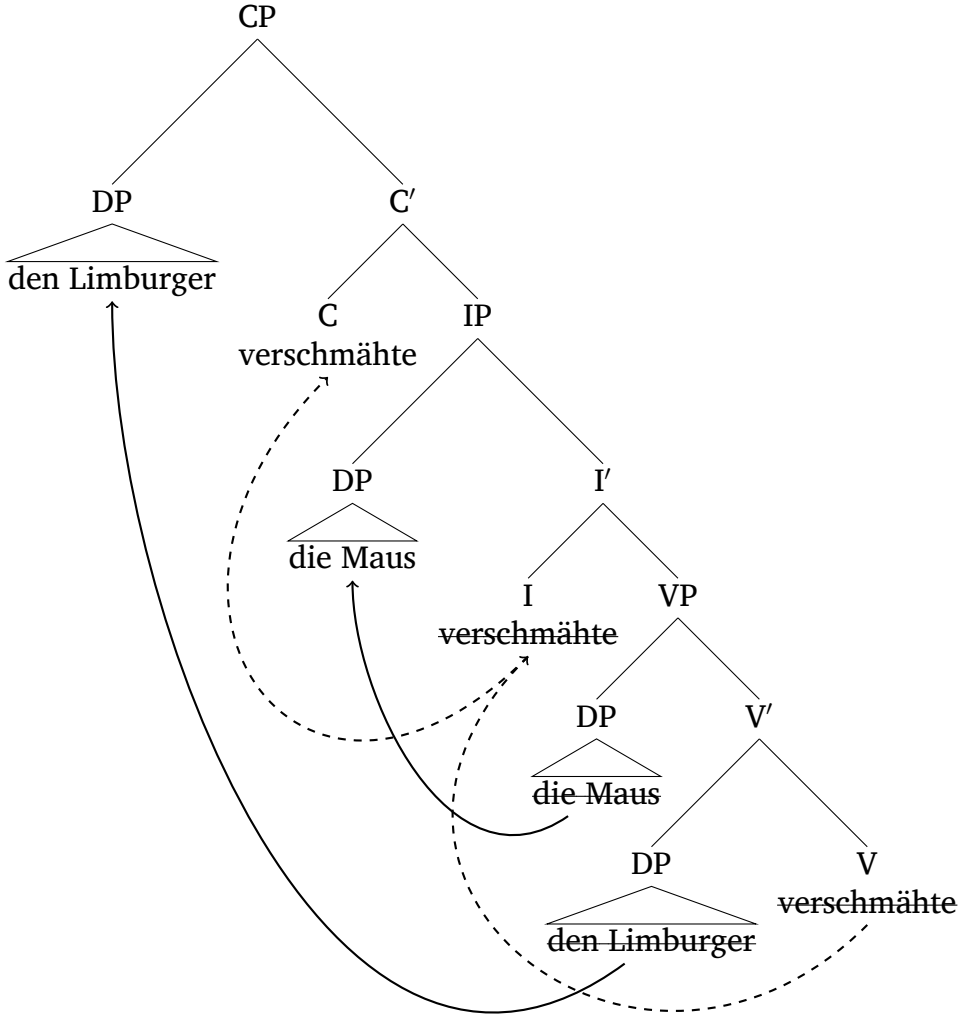
[14] The assumption that C bears [TENSE] derives from the observation that complementisers select either finite or non-finite VPs. In English for instance, *for* combines with to-infinitives, whereas *that* and *if* combine with finite VPs (den Besten 1989: 90). Furthermore, in certain Dutch and German varieties, complementisers agree with finite verbs (den Besten 1989: 92), cf. the Hollandic example in (i).

(i) …datte-e ze komme-e HOLLANDIC
…that-PL they come-PL
(den Besten 1989: 93)

[15] See also Frey (2006a) and Light (2012) for a proposal in which the clause-initial position can be filled via information-structurally motivated Ā-movement or via so-called formal movement to simply satisfy an [EPP]-feature.

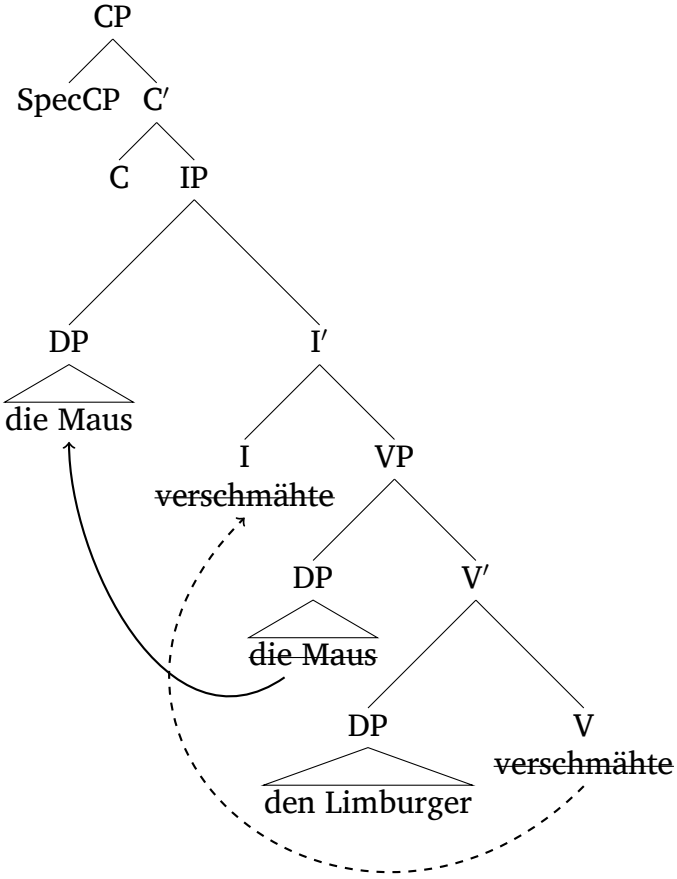
(31) a. Den Limburger **verschmähte** *die Maus*. GERMAN
the.ACC limburger cheese spurned the.NOM mouse
'The mouse spurned the limburger cheese.'

b.



(32) a. *Die Maus* **verschmähte** den Limburger. GERMAN
the.NOM mouse spurned the.ACC limburger cheese
'The mouse spurned the limburger cheese.'

b.



Empirical support for asymmetrical analyses comes again from the placement of weak pronouns and clitics in German and Dutch. Consider the contrast between the placement of subjects and objects in (33) and (34). Nominal subjects (33a) and subjects represented by weak pronouns (33b) can both appear clause-initially. However, whilst no restrictions apply to clause-initial nominal objects (34a), weak object pronouns (34b) are barred from the clause-initial position according to Travis (1991: 359) and Zwart (1997: 35, 196). This pattern directly follows if it is assumed that weak pronouns are somehow barred from SpecCP: Weak subject pronouns can only appear clause-initially, because they are realised in a lower position.[16]

(33) a. *Das Mäuschen* **hat** das Bündnerfleisch beschnuppert. GERMAN
the mouse.DIM has the Grisons meat sniffed
'The little mouse sniffed at the Grisons meat.'

[16] Travis (1991: 359) speculates that only pronouns with the ability to bear focal stress may be moved to SpecCP.

b. *Es* **hat** das Bündnerfleisch beschnuppert.
it has the Grisons meat sniffed
'It sniffed at the Grisons meat.'

(34) a. Das Bündnerfleisch **haben** *die Mäuse* beschnuppert. GERMAN
the Grisons meat have the mice sniffed
'The mice sniffed at the Grisons meat.'
b. *Es **haben** *die Mäuse* beschnuppert.
it have the mice sniffed
'The mice sniffed at it.'

However, for German at least a generalised ban on clause-initial weak pronominal objects (i.e. *es* 'it') is not empirically supported (Frey 2006b, Meinunger 2007). Sentences with preposed object *es* become in fact acceptable when the subject remains in a low position (35) (Meinunger 2007).[17,18]

(35) Q: Warum habt ihr das Huhn immer noch? GERMAN
why have you the chicken always still
'How come you still have the chicken?'
A: Es **konnte** einfach *niemand* hier schlachten.
it could simply nobody here slaughter
'Well, nobody here was able to kill it.'
(Meinunger 2007: 557)

According to Zwart (1997: 195), further evidence for different positions of initial subjects and non-subjects can be found in certain varieties of Dutch with a specific type of complementiser agreement. In varieties with what Zwart (1997: 138) dubs *double agreement*, two paradigms of agreement markers can be found: one for complementisers and one for verbs. In East Netherlandic, Brabantish and West Flemish, finite verbs in embedded clauses (36a) and in subject-initial main clauses (36b) bear verbal agreement. Finite verbs in non-subject-initial sentences (36c), however, are marked with the complementiser agreement forms (Zwart 1997: 139). The distribution of different agreement markers can be directly accounted for if different agreement markers correspond to different positions (i.e. CP → complementiser agreement, IP → verbal agreement).

(36) a. …datte wy speul-t/*-e EAST NETHERLANDIC
…that.1PL(C) we play-1PL(V)/-1PL(C)
b. Wy speul-t/*-e
we play-1PL(V)/-1PL(C)
c. Waor speul*-t/-e wy?
where play-1PL(V)/-1PL(C) we
'Where do we play?'
(Zwart 1997: 140)

[17] Albeit not a prerequisite, unspecific indefinite subjects — as in (35) — further enhance the acceptability of such constructions (Meinunger 2007: 557).
[18] See Schwartz & Vikner (1989, 1996) for further counterarguments.

In addition to the empirical evidence outlined above, asymmetrical accounts avoid conceptual problems that symmetrical analyses face. Both Travis (1991: 361) and Holmberg (2015: 364) point to potential learnability problems for children (cf. also Westergaard 2009b): The movement of subjects and verbs from IP to CP is string-vacuous as the movement yields the same configuration as before the movement. This poses the question how learners would be able to identify movement to CP in the input data without clear evidence. That is, symmetrical accounts cannot straightforwardly explain how children acquire the movement of subjects and verbs to the CP-domain without any additional stipulations.

1.2.3.3 Articulated left periphery analyses

The final type of V2 analysis sketched here assumes a complex left periphery with multiple projections instead of a single CP projection. It has already been recognised in symmetrical accounts that a simple CP structure may not be sufficient for capturing all V2 languages. Unlike Dutch and German — where embedded V2 can only occur in the absence of a complementiser (cf. (28)) — Mainland Scandinavian languages allow embedded V2 with an overt complementiser (Wiklund et al. 2009, Heycock, Sorace & Hansen 2010, Julien 2015, Westendorp & Lundquist 2022[19]).[20] The example in (37) shows this pattern clearly.

(37) [Johan sa] att igår **kom** *Ida* inte for sent till skolan. SWEDISH
Johan said that yesterday came Ida not too late to the.school
'Johan said that Ida didn't come late for school yesterday.'
(Westendorp & Lundquist 2022: 150)

To explain the co-occurrence of complementisers and V2, some symmetrical analyses have proposed a recursive CP (Iatridou & Kroch 1992, Vikner 1995, cf. also Branigan 1996). The structure of the Swedish example in (37) can thus be analysed with two CPs, that is a lower and a higher CP: The complementiser resides in the higher CP, whereas the finite verb and the preceding constituent occupy the lower CP.[21]

[19]The manuscript of Westendorp & Lundquist (2022) was first published as chapter 4 in Westendorp (2022).

[20]The situation in German might also not be as clear-cut as previously assumed. Freywald (2008, 2009, 2016) has noted the existence of embedded V2 sentences with an overt complementiser *dass* 'that' in spoken German, illustrated in (i). According to Freywald (2008, 2009), *dass*-V2 constructions are associated with specific pragmatic properties expressing assertion. It remains unclear though whether this construction is subject to any geographical restrictions. At least in the Southern German variety spoken by the author of this thesis, *dass*-V2 does not constitute a grammatical construction.

(i) Ich habe gelesen, dass in Sizilien **gibt**'*s* welche, die sind 'n paar hundert Jahre alt.
I have read that in Sicily give=it some they are a couple hundred years old
'I've read that there are some on Sicily, they are a couple of hundred years old.'
(Freywald 2016: 335)

See also de Haan (2001) for similar observations in Frisian.

[21]Though see Westendorp & Lundquist (2022) for an analysis of the Mainland Scandinavian languages with a single CP.

The idea of multiple positions in the left periphery has also been adopted by analyses assuming a complex articulated left periphery. Specifically, the CP is split into several distinct projections, following seminal work by Rizzi (1997). Two broad categories of projections are generally distinguished: One the one hand, Fin and Force which relate to the 'propositional content' and higher structure (e.g. discourse), respectively (Rizzi 1997: 283). On the other hand, projections that encode information-structural notions such as focus and topic. The exact hierarchical composition of the left periphery varies between proposals. Holmberg (2015), for instance, provides the hierarchy in (38), based on Poletto (2002).

(38) [Hanging Topic [Scene-setting [Force [Topic [Focus [WH [Fin]]]]]]]
(Holmberg 2015: 373)

In analyses assuming a complex left periphery, the finite verb moves to one of the projections in (38), albeit the exact position varies between theories (e.g. Poletto 2000, 2002, Westergaard & Vangsnes 2005, Frascarelli & Hinterhölzl 2007, Mohr 2009, Walkden 2014). A major advantage of this strand of analysis is that relaxed V2 languages (such as the Medieval Romance languages, cf. §1.2.2 and §1.3) where V2 and V>2 orders alternate can readily be explained. Due to the articulated left periphery multiple landing sites are available which can host the various preverbal constituents. As a result of this type of analyses (but also due to the symmetrical analyses), the linear definition of V2 is usually discarded in favour of a more technical definition: V2 is the movement of the finite verb to a head in the CP-domain combined with movement of a XP to a specifier in the left periphery (Mohr 2009: 156, Cognola 2015, 2019, Holmberg 2015: 375, Wolfe 2019: 36). To limit the amount of moved constituents to the CP-domain to one, many authors posit that all constituents need to move to SpecFinP first in order to access the left periphery (Poletto 2002, Roberts 2004, Mohr 2009). The copy (or trace) left behind after the constituent is moved to a higher position then prevents the movement of any further constituents to the CP-domain. In other words, SpecFinP acts as bottleneck.[22] Having established the theoretical background, I can now turn to the diachrony of V2.

[22]There is more to say about the positions targeted by the clause-initial constituent and the finite verb in the CP-domain. For instance, Wolfe (2019) distinguishes between Fin-V2 and Force-V2. This distinction has implications for the analysis of relaxed V2-languages: In a Fin-V2 system positions for multiple preverbal constituents are available, whereas a Force-V2 system can merely host one additional preverbal constituent (i.e. a frame-setter) in the structure suggested by Wolfe (2019). That is, different types of relaxed V2 languages can be distinguished: those that allow only one type of additional preverbal constituent (Force-V2) or those that allow multiple types (Fin-V2). Note that SpecFinP still acts as bottleneck in that the other preverbal constituents are assumed to be base-generated in their left-peripheral position (Wolfe 2019). A more detailed discussion would, however, lie outside the scope of this thesis.

1.3 The loss of V2

As noted in §1.2.2, some languages which had been characterised by a V2 grammar at some point in their history subsequently lost this property. Languages falling into this group are English, most of the Romance languages and Welsh. This group of languages stands in stark contrast to other V2 languages which have been characterised by a continuity of relatively strict V2 for the entirety of their written records. Some of the languages in the latter group (which encompasses most of the Germanic languages) even developed a stricter V2 system. In Old High German (OHG), for instance, the movement of the finite verb had almost completely been generalised, whereas XP-fronting was more variable (Axel 2007, 2009a,b, Axel-Tober 2015, 2018, Cichosz 2010): On the one hand, XP-movement could be completely absent given the existence of V1 sentences, such as the existential construction in (39a) from Tatian. The present-day High German (PDG) equivalent in (39b), in contrast, is ungrammatical without the insertion of the expletive *es* 'it' in clause-initial position. On the other hand, cases of V > 2 are also attested. In Isidor for example, instances of V3 sentences can be found where a pronoun intervenes between a fronted XP and the finite verb (40a). The Modern German equivalent in (40b) is again ungrammatical.

(39) a. **uuas** thar ouh sum uuitua in thero burgi OHG
was there also some widow in that city
'There was a certain widow in the same city …'
(Axel-Tober 2018: 41)

b. *(Es) **gab** eine bestimmte Witwe in der gleichen Stadt PDG
EXPL gives a certain widow in the same city

(40) a. Erino portun *ih* **firchnussu** OHG
iron portals I.NOM destroy
'I will destroy iron portals'
(Axel-Tober 2018: 32)

b. *Eisentore *ich* **werde** zertstören PDG
iron portals I will destroy

This development is not unique to German but has been observed for other Germanic languages as well, such as Old Saxon and Old Norse. Both exhibited generalised V-to-C movement (Eythórsson 1995: 189, Þorgeirsson 2012: 234), but variation in the realisation of XP-movement (Faarlund 2004: 191–192, Walkden 2014: 92–93).

These two diverging diachronic developments pose the question why some languages 'reinforce' their V2 grammars whilst others 'dispose' of it. Do the changes in individual languages constitute unique developments in that the contributing factors are specific to each language? Alternatively, can parallels be observed between different languages? That is, can the loss of V2 be attributed to a set of shared factors? In this section, I will summarise different accounts for the loss of V2 in English (§1.3.1), French (§1.3.2), Portuguese (§1.3.3) and Welsh (§1.3.4). This overview will show that the factors leading to the loss of V2 are language-specific, but the trajectory of the loss of V2 shares important similarities, mostly in the distribution of clause-initial constituents (§1.3.5).

1.3.1 English

To understand the changes that have occurred in the history of English, it is necessary to sketch the prevalent word order patterns first, in particular with respect to verb placement. The word order in Old English (OE) is comparable to the order noted above for OHG and other early Germanic languages (Hinterhölzl & van Kemenade 2012). That is, unlike present-day German, OE is not a strict V2 language and word order variation between V1 (41), V2 (42)–(43) and V3 sentences (44) has been reported (Walkden 2014).[23]

(41) **Wæs** *he se biscop* æfest mon & god OE
was he the bishop pious man & good
'He the bishop was a pious and good man.'
(Walkden 2014: 92)

(42) a. ne **sceal** *he* naht unaliefedes don OE
not shall he nothing unlawful do
'He shall not do anything unlawful.'
b. þa **wæs** *þæt folc* þæs micclan welan ungemetlice brucende …
then was the people the great propserity excessively partaking …
'Then the people were partaking excessively of the great prosperity.'
(Fischer et al. 2001: 106)

(43) a. Him **geaf** ða *se cyngc* taw hund gildenra pænega OE
him gave then the king two hundred golden pennies
'Then, the king gave him two hundred pence in gold.'
b. On his dagum **sende** *Gregorius* us fulluht
in his days sent Gregory us baptism
'In his time, Gregory sent us baptism'
(Haeberli 2002a: 88)

(44) Be ðæm *we* **magon** suiðe swutule oncnawan ðæt … OE
by that we may very clearly perceive that …
'By that, we may perceive very clearly that …'
(van Kemenade & Westergaard 2012: 94)

The word order variation illustrated in (41)–(44) is contingent on different factors. If a clause forms a *wh*-question or if either the negative adverb *ne* or the adverbials *þa/þonne* 'then' are realised clause-initially (42), the V2 rule is strictly obeyed (Fischer et al. 2001, van Kemenade & Westergaard 2012, Haeberli, Pintzuk & Taylor 2020).[24] In

[23] Note though that this represents only a partial and simplified picture as other word orders have been described in the literature. For instance, verb-late or verb-final clauses are attested in OE (Fischer et al. 2001, Pintzuk & Haeberli 2008, Cichosz 2010, Walkden 2014).

[24] Due to the mandatory nature of V2 following *þa/þonne* 'then', the adverbial is often treated as operator(-like) in the literature, on a par with *ne* and *wh*-elements (van Kemenade 1987, Pintzuk 1991, Kroch & Taylor 1997). Westergaard (2009d), however, rightly points out that such a characterisation is problematic because the equivalents of *then* in other languages "are not universally operators" (Westergaard 2009d: 93).

other contexts, the nature of the subject seems to determine the word order: Nominal subjects follow the verb (43), whilst pronominal subjects generally appear preverbally (44) (Fischer et al. 2001, van Kemenade & Westergaard 2012, Haeberli, Pintzuk & Taylor 2020).[25] Although pronominal subjects account for the majority of the elements left-adjacent to the finite verb in V3 constructions, nominal subjects do in fact occur in this position as well (45). Based on a sample from ten OE texts, Haeberli (2002b: 250) concludes that the absence of inversion with nominal subjects is non-negligible with almost 30% on average (albeit significant variation between texts was observed). According to Kroch, Taylor & Ringe (2000: 365), this configuration is particularly frequent with a clause-initial scene-setting temporal.

(45) æfter þan *þæt lond* **wearð** nemned Natan leaga OE
after that that land was named Natan lea
'After him that land was called Netley.'
(Haeberli 2002b: 249)

To account for the alternation between V2 and V3, it has been argued that the position of subjects is conditioned on information-structural factors: Subjects are realised in the position preceding the finite verb, if they are given by the preceding discourse. On the other hand, if subjects constitute new (i.e. non-given) information, they are placed in the position following the verb (Bech 2001, Westergaard 2009d, Biberauer & van Kemenade 2011, van Kemenade & Melićev 2011, van Kemenade 2012, van Kemenade & Westergaard 2012, Speyer 2010, 2012). Moreover, the clause-initial position is also subject to information-structural properties in that it links the proposition to the preceding discourse (van Kemenade & Los 2006, Los 2009, 2012, Los & Dreschler 2012, Los & van Kemenade 2018). Based on those findings, Haeberli, Pintzuk & Taylor (2020) suggest the structure in (46) for OE. The finite verb targets CP2 in sentences with strict V2 (i.e. *wh*-questions and sentences headed by *ne* and *þa/þonne*), while CP1 is the target in all other contexts. As for given subjects, they reside in SpecCP1 (=SU1). Non-given subjects, on the other hand occupy SpecTP (=SU2). This structural analysis has not been unanimously accepted in the literature though. Fischer et al. (2001), among others, posit that CP1 corresponds to an (otherwise unspecified) functional projection couched between the CP- and IP-domain, namely FP (47).

(46) $[_{CP2}$ XP C2 … $[_{CP1}$ SU1(pro/DP) C1 $[_{TP}$ SU2(DP) T]]]
(Haeberli, Pintzuk & Taylor 2020: 399 following Walkden 2017a)

[25] Pronominal subjects may also co-occur with pronominal objects (Pintzuk 1991, Fischer et al. 2001). In the example in (i), both a pronominal subject and pronominal indirect object precede the verb in addition to a nominal object.

(i) & seofon ærendracan *he* him **hæfde** to asend OE
& seven messengers he them had to sent
'…and he had sent seven messengers to them.'
(Pintzuk 1991: 188)

	pronominal subjects				nominal subjects			
	M1	*M2*	*M3*	*M4*	*M1*	*M2*	*M3*	*M4*
Aux	27.9	26.5	33.4	30.6	74.5	53.4	44.1	48.9
Trans + Unerg	23.7	10.3	12.8	12.2	66.8	51.7	31.2	22.9
Unacc	26.9	11.8	15.3	17.2	70.2	52.6	54.4	57.2

Table 1.1: Percentage of subject-verb inversion with pronominal and nominal subjects in ME (van Kemenade & Westergaard 2012: 100, 102). M1 to M4 represent the following time periods: M1 = 1150–1250, M2 = 1250–1350, M3 = 1350–1420, M4 = 1420–1500. The percentages were calculated for different classes of verbs, that is auxiliaries (aux), transitive and intransitive/unergative verbs (trans + unerg) as well unaccusatives (unacc).

(47) $[_{CP}$ XP C $[_{FP}$ SU1(pro/DP) F $[_{TP}$ SU2(DP) T]]]
(Fischer et al. 2001: 126)

The transition from OE to Middle English (ME) did not see significant changes initially, at least in terms of the V2 system (van Kemenade 1997, Haeberli 2002b: 252). In later ME, however, two (coarse) varieties can be distinguished (Kroch & Taylor 1997, Kroch, Taylor & Ringe 2000). One variety, spoken in the south of England, is characterised by the continuation of the OE pattern. The second variety, used in the north, is more similar to the modern West Germanic and Mainland Scandinavian languages: Both nominal and pronominal subjects typically invert if a constituent other than the subject occupies the clause-initial position (Kroch, Taylor & Ringe 2000: 372).[26,27] The upheaval of the V2 grammar began towards the end of the 14th century and the beginning of the 15th century, when a series of changes commenced: First, the frequency of topicalised nominal objects decreased significantly (Speyer 2010: 66, van Kemenade & Westergaard 2012: 104) after topicalised pronominal objects had already declined a century prior (Haeberli, Pintzuk & Taylor 2020). Second, the proportion of V2 sentences with clause-initial PPs started to dwindle (Speyer 2010: 65). The last change that occurred was the loss of subject-verb inversion with nominal subjects (van Kemenade 1997, Fischer et al. 2001), although some qualifications are in order, as van Kemenade & Westergaard (2012) observe: The proportion of postverbal nominal subjects decreased continuously throughout the ME period (Table 1.1). At the time of the other changes, however, the proportion of postverbal subjects actually started to increase again with auxiliaries (and to some extent with unaccusative verbs). Interestingly, the same pattern can be observed for pronominal subjects. That is, the amount of inverted pronominal subjects increased in sentences with finite auxiliaries (van Kemenade & Westergaard 2012: 100).

[26] Kroch & Taylor (1997) and Kroch, Taylor & Ringe (2000) claim that language contact with Scandinavian invaders can be made responsible for the emergence of the northern variety. See Walkden (2021b, to appear) for a critical discussion of this claim.

[27] See also Truswell (2021) for compelling arguments that at least one further Northern variety exists.

At the end of the ME period however, this trend was reversed and the frequency of subject-verb inversion was falling again. At the end of these changes, the V2 grammar had to give way to the SVO grammar found in Modern English.[28]

As briefly mentioned above, the syntactic structure of OE and early ME has been subject to debate such that no analysis has been unamiously accepted. One of the main points of contention is the exact location of SU1 in the tree (cf. (46) vs. (47)). The ramifications of different position are non-trivial: Depending on the exact positions, the changes that occurred in the history of English may either be considered conservative or more profound. For instance, a proposal falling into the first category has been developed by Fuss (2003, 2008). Contrary to Haeberli, Pintzuk & Taylor's (2020) analysis that SU1 is located in the left periphery, Fuss (2003, 2008) argues that SU1 corresponds to SpecTP and SU2 to Spec*v*P. This entails, if the standard analysis of English is assumed, that only the position of nominal subjects has changed; pronominal subjects have resided in SpecTP ever since the OE period.[29] Fuss (2003) attributes the changed position of nominal subjects to the emergence of an [EPP]-feature on T.[30] Due to the presence of the EPP-feature, SpecTP has to be overtly filled, triggering the movement of nominal subjects from *v*P. As for the rise of the [EPP]-feature, Fuss (2003: 221) speculates that the evolution of a tense system with clear functional partition might be the driving factor. In either case, the analysis of Fuss (2003, 2008) suggests that only a 'pseudo-V2' system was lost.

A different position is advocated by Haeberli (2002a) who ties the loss of V2 to changes in the inflectional domain. Importantly, the demise of V2 is not argued to be the immediate result of changes to the inflectional morphology; rather, the morphological changes affected other properties of the grammar that then lead to the loss of V2. According to Haeberli (2002a), the property connecting the loss of V2 to changes in the morphology are empty expletives (i.e. *pro*). For empty expletives to be licensed by the grammar, the verbal agreement paradigm must not exhibit any syncretisms between the infinitival marker(s) and singular forms (Haeberli 2002a: 100). This is based on observations from the modern West Germanic languages where West Flemish is set apart from the the other languages by the ungrammaticality of expletive *pro* (Haeberli 1999) — exemplified in (48). Crucially, West Flemish is also the sole language (among the West Germanic languages) in which the markers for the first person singular (*-en*) and the infinitive (*-en*) are synrectic (Haeberli 2002a: 100 p.c. Liliane Haegeman).

(48) a. …dass *pro* überall getanzt **wurde** GERMAN
…that *pro* everywhere danced was
'…that people danced everywhere'

[28]Note that the loss of verb movement occurred independently from the loss of V2 (Haeberli & Ihsane 2016).

[29]See also Kiparsky (1995) for arguments that the changes in OE are more conservative than in other Germanic languages when considered from the perspective of other Indo-European languages.

[30]According to Fuss (2003), pronominal subjects have to undergo overt fronting in OE because T bears uninterpretable case and φ-features. This set of features can trigger overt movement of formal feature bundles (i.e. pronouns).

	OE	Early ME	Late ME
INF	*-an*	*-en*	*-e*
1SG	*-e*	*-e*	*-e*
2SG	*-st*	*-st*	*-st*
3SG	*-þ*	*-þ*	*-þ*

Table 1.2: Verbal agreement in Old English (OE) and Middle English (ME) for singular and infinitival forms (Haeberli 2002a: 101). The agreement morphology remains mostly unchanged with the notable exception of the infinitive.

b. …dat *pro* overal gedanst **werd** DUTCH
…that *pro* everywhere danced was
c. …dat *(er) overal gedanst **wier** WEST FLEMISH
…that there everywhere danced was
(Haeberli 2002a: 96)

To explain the loss of V2 in English, Haeberli (2002a) adopts an analysis originally proposed by Haeberli (1999) for V2 sentences in contemporary West Germanic languages. Recall from §1.2.3.1 that whilst weak pronouns need to be right-adjacent to the finite verb in Dutch, no such constraint applies to strong pronouns. Nominal subjects pattern with strong subject pronouns in that they are exempt from this adjacency constraint. This is not unique to Dutch but has been noted for other Modern West Germanic languages as well. West Flemish again constitutes the only exception as nominal subjects need to be left-adjacent to finite verbs (Haeberli 1999). Given the correlation between empty expletives and the adjacency constraint in different West Germanic languages, Haeberli (1999) suggests that nominal subjects can remain in a lower position than pronominal subjects because an empty expletive is merged in the higher position.[31] The absence of expletive *pro* in West Flemish can thus explain why both nominal and pronominal subjects need to be realised adjacent to the finite verb.

Returning to English, Haeberli (2002a) argues that this analysis can also be applied to non-operator contexts in OE. That is, subjects in OE and the Modern West Germanic languages are structurally identical: Pronominal subjects are moved to SU1, whereas nominal subjects typically remain in SU2 (cf. (47)). This entails that the V2 order of OE in non-operator contexts can be derived from the presence of an empty expletive in sentences with non-clause-initial nominal subjects, thus allowing the nominal subjects to remain in the lower position (Haeberli 2002a).[32] The loss of expletive *pro* would thus force the movement of nominal subjects to the higher SU1 position, eventually resulting in non-V2 configurations. Table 1.2 illustrates the development of the verbal

[31] In order to explain the obligatory high position of pronominal subjects, Haeberli (2002a: 96) hypothesises that pronominal subjects need to move to the higher position due to some 'licensing requirement'.
[32] The existence of expletive *pro* has been independently argued for by Hulk & van Kemenade (1995).

agreement markers in the singular in the history of English. While the infinitive does not show any syncretisms with singular agreement markers in OE and early ME, this picture changes in late ME. Due to the loss of final /n/ in infinitives, the first person singular and the infinitive bear the same marker, namely *-e*. Based on the established licensing condition for empty expletives in West Germanic languages, Haeberli (2002a) concludes that expletive *pro* is no longer licensed in late ME due to the syncretism illustrated in Table 1.2. Consequently, V2 was no longer possible in non-subject-initial sentences as nominal subjects could no longer remain in SU2. The connection between expletive *pro* and V2 is further supported by the fact that the demise of both fall into the same period (Hulk & van Kemenade 1995: 249).[33] Moreover, the presence of /n/ in the infinitival marker and S-V inversion are positively correlated in various ME texts (Haeberli 2002b). Taken together, a simple change in another area of the grammar may have had significant ramifications in the syntax.[34]

A different explanation for the loss of V2 has been suggested by van Kemenade & Westergaard (2012). Unlike the previous explanations, van Kemenade & Westergaard (2012) do not attribute the loss of V2 to morphological changes — instead, the aforementioned information-structural reorganisation in ME is identified as one of the driving factors. According to van Kemenade & Westergaard (2012), the first necessary development towards the loss of V2 was the rise of preverbal subjects.[35] This development is attributed to the fact that subjects tend to represent given information and are thus often realised as pronouns, especially in spoken language (van Kemenade & Westergaard 2012). In fact, Westergaard (2010) found that over 90% of all subjects are realised as pronouns in spoken language, irrespective of the clause type. For ME, this tendency meant a preponderance of (XP)SV structures as given subjects were normally placed before the verb. The prevailing lack of subject-verb inversion encountered by ME speakers eventually caused some new learners to posit a default (XP)SV order.[36,37] The interim increase of V2 structures with auxiliaries, in turn, follows

[33] One may ask, however, whether the mere concurrence of the loss of expletives and the V2 suffices as evidence. The analysis of Haeberli (2002a) crucially requires that the advent of the loss of expletive *pro* predates the one of the loss of V2.

[34] A potential problem for Haeberli's (1999, 2002a) proposal arises due to the prediction of V1 sentences. If no other constituent is fronted to the clause-initial position and the enumeration contains an expletive *pro*, the nominal subject should be able to remain in the lower position (*pro*-V-S). Although V1 sentences are indeed attested in OE, V1 sentences occur in a very restricted environment in the modern West Germanic languages (for German e.g. Önnerfors 1997, Freywald 2013). One would need to stipulate that *pro* cannot be clause-initial or that the clause-initial position needs to be phonologically filled.

[35] The importance of preverbal subjects has also been recognised by van Kemenade (2012). In fact, the increasing number of preverbal subjects is identified as the most important change by van Kemenade (2012: 822).

[36] The development of a default preverbal subject position was also observed in interrogatives and subordinate clauses (van Kemenade & Westergaard 2012: 112).

[37] van Kemenade & Westergaard (2012: 112) briefly allude to the actuation problem (cf. Walkden 2017b). That is, why does the change happen during the ME period and not earlier given that the position of subjects was also information-structurally conditioned in OE? van Kemenade & Westergaard (2012) claim that different factors need to conspire for changes to spread among speakers. That is, not all prerequisites were met in earlier stages of English.

from the co-existence of speakers using the older OE-like grammar and speakers using the new grammar with default high-subject position (van Kemenade & Westergaard 2012: 113). Children acquiring language in this community would be unable to find clear information-structurally-conditioned patterns due to the competition between the two grammars (cf. Kroch 1989). Instead, learners would interpret the information-structurally-conditioned input as syntactically conditioned V-to-C movement, albeit only with auxiliaries (van Kemenade & Westergaard 2012: 113).[38] Later generations would then have faced an even more confusing picture, rendering the grammar unstable. In the face of an unstable grammar, economy principles that are active during language acquisition would have led to the eventual loss of V2 (van Kemenade & Westergaard 2012: 114); the exerted pressures to simplify the complex grammar would lead learners to assume a more homogeneous SVO grammar.

A final account for the loss of V2 in English sketched here is sociolinguistic in nature. As noted above, two (broad) dialects of ME have been distinguished based on their syntax — a northern variety and a southern variety. The southern variety exhibits similar word order patterns as OE. The northern variety, on the other hand, resembles modern Germanic V2 languages more closely considering its more rigid V2 syntax. Kroch, Taylor & Ringe (2000) state that the co-existence of a northern and a southern dialect led to a situation of grammar competition in the speaker community due to language contact (cf. Kroch 1989, 1994). When interacting with speakers from the south, "northern speakers would try to accommodate […] their interlocutors" (Kroch, Taylor & Ringe 2000: 377). However, the lack of subject-verb inversion in constructions with pronominal subjects as well as in constructions with clause-initial scene-setters (both with nominal and pronominal subjects) would lead northern speakers to posit a non-V2 grammar and mix it with their V2 grammar.[39] The resulting mixed language would be learnt by children and would spread into the South. Over successive generations, the proportion of the non-V2 grammar would then steadily increase, resulting in the elimination of the V2 grammar (Kroch, Taylor & Ringe 2000: 377, cf. Yang 2000, 2002). Haeberli (2002a: 93) notes that such an account faces considerable issues: Finding evidence that either supports or refutes such an analysis may not be possible. I will briefly return to this issue later in §1.3.5. In the next section, I will discuss the loss of V2 in French.

[38] Some questions remain unanswered with this analysis. First, what learning trajectory do learners follow during language acquisition? Is learners' initial stipulation to search for information-structurally-conditioned rules? Second, why do learner restrict the V-to-C analysis to auxiliaries? Arguably, there must have been an increased occurrence of V2 with auxiliaries, otherwise learners would not have derived such a rule.

[39] Kroch, Taylor & Ringe (2000: 377) claim that a non-V2 grammar must be the unmarked option, given the typological scarcity of V2 (cf. §1.2.2).

1.3.2 French

The word order in Old French (OF) has attracted significant attention among scholars.[40] Many have argued that OF constitutes a V2 language (Roberts 1993, Vance 1995, Wolfe 2018c). This analysis is motivated by the observation that different types of constituents can appear in the clause-initial position, including subjects (49a), direct objects (49b), adverbs (49c) and verbal complements (49d) (Adams 1987a: 26, Wolfe 2018c: 67, Larrivée 2021: 191). Crucially, subject and verb typically invert if the clause-initial position is hosting a non-subject, as the examples in (49b)–(49d) amply illustrate.

(49) a. *Il* **oste** ses armes OF
he remove.3SG his weapons
'He removes his weapons.'

b. son cors ne **poï** *je* veoir
his body NEG can.1SG I see.INF
'I cannot see his body.'

c. Longuement **parlerent** ensemble entre *le preudome et Lancelot*
long speak.3PL.PST together between the nobleman and Lancelot
'The nobleman and Lancelot spoke together for a long time …'

d. Chanceler te **fist** *il*
shake.INF you.CL make.3SG.PST he
'He made you shake.'

(Wolfe 2018c: 67–68)

The exact information-structural composition of the clause-initial position has been contended. Even though a consensus has been reached that clause-initial constituents can bear different informational values (including none) (Steiner 2014, Wolfe 2018c), the concrete composition has been subject to debate. While Labelle & Hirschbühler (2018) argue that until the 13th century a majority of all clause-initial objects constitute foci, Ingham (2018: 249) and Larrivée (2019) found a large proportion of topic-initial sentences. Larrivée (2022) attributes these diverging findings to differences in register. In the sample studied by Larrivée (2022), literary texts are much more likely to be focus-initial than legal texts. Alternatively, the difference may also be attributable to the substantial microvariation noted for OF (Wolfe 2018b). That is, different regions may manifest different preferences for the informational values of clause-initial constituents. These explanations are not mutually exclusive and both might have contributed to the diverging observations.

Similar to English, deviations from the V2 order have been observed in OF. Bech & Salvesen (2014) remark, however, that the deviations in OF are considerably more homogeneous than in OE. In principle, two types of deviations have been noted in the literature: V1 and V3 clauses. The first type, i.e. V1 clauses as exemplified in (50), have been argued to be rare and even almost completely absent in the 13th century (Wolfe 2018c: 75). This claim was not borne out in the studies of Kaiser & Zimmermann (2011),

[40] The French spoken until the 13th century is typically labelled as Old French (Vance 1997: 1), whereas the French of the 14th and 16th century is referred to as Middle French (Wolfe 2021: 2).

Century	V1	V2	V>2
13th	.10	.72	.18
14th	.17	.55	.28
15th	.15	.54	.31
16th	.12	.54	.34

Table 1.3: Proportion of different word orders in Medieval French main clauses by century (Steiner 2014: 129). The proportion of V1 and V>2 sentences increases from the 13th to the 14th century. This comes at the cost of V2 sentences whose proportion significantly declines during this period. Subsequent time periods saw a more stable proportion of different word orders.

Zimmermann (2014) and Steiner (2014). In the scrutinised sample of Steiner (2014), for instance, V1 sentences account for 10% of all structures in the 13th century (cf. Table 1.3). It should be pointed out though that OF is a null subject language (Adams 1987a,b, Roberts 1993, Vance 1997). Consequently, V1 sentences such as the one in (51) could also be analysed as [*pro* V] structures.

(50) **Respundi** *Samuel*: 'cument'? OF
respond.3SG.PST Samuel how
'Samuel responded: "how?"
(Wolfe 2018c: 75)

(51) **Vint** en Bethléém OF
come.3SG.PST in Bethlehem
'He came to Bethlehem.'
(Wolfe 2018c: 75)

V3 clauses, i.e. the second type of deviating word order pattern, on the other hand, occur more frequently than V1 sentences in OF (cf. Table 1.3; Kaiser & Zimmermann 2011: 363, Steiner 2014: 129). The composition of V3 sentences follows mainly two types: Either an adverbial or clausal frame-setter (i.e. elements setting the frame for the conveyed information) precedes the verb and the preverbal constituent (52a) or a hanging topic that is resumed by a resumptive pronoun (52b) (Wolfe 2018c: 75).

(52) a. Et neporec *Nostre Sires* **avoit** mis … OF
and nevertheless our Lord have.3SG put.PTCP …
b. Li chevalier qui sont en pechié mortel, *ce* **sont** …
the knights that be.3SG in sin mortal they be.3PL …
(Wolfe 2018c: 76)

The transition from OF to Middle French (MF, cf. fn.40) is characterised by significant changes in the distribution of different word order patterns. Steiner (2014) scrutinised

Century	SV	VS	NS
13th	.47	.23	.30
14th	.57	.20	.23
15th	.64	.10	.26
16th	.63	.06	.31

Table 1.4: Proportion of subject-initial (SV), non-subject-initial (VS) and null subject (NS) V2 sentences in Medieval French. The values reported here were calculated based on the total number of V2 sentences and different types of word orders provided by Steiner (2014: 129). While the proportion of subject-initial V2 sentences increases, the proportion of non-subject-initial sentences decreases at the same time. V2 sentences with null subjects first decrease but return to their initial proportion in the 16th century.

the developments of V1, V2 and V3 clauses from the 13th to the 16th century in a sample of 600 sentences per century. The results are reproduced in Table 1.3. The study revealed that the proportion of V3 clauses increased noticeably from the 13th to the 14th century. This coincides with two further developments. First, the amount of preverbal subjects increased. This becomes evident when the proportion of different clause-initial constituents in V2 sentences in Table 1.4 is considered: The proportion of subject-initial sentences increases by 10% in the 14th century compared to the previous century. Second, the composition of fame-setters shifted from predominantly clausal frame-setters to PPs and adverbial frame-setters. This trend continues into the 15th century. While V3 sentences followed predominately the same pattern in the 14th century, the 15th century sees the rise of constructions of the type Focus-XP-V. In the 16th century previous changes further solidified. That is, subjects became increasingly realised before the verb. The Focus-XP-V construction, however, declined again. Given the low frequency of postverbal subjects and the high frequency of V3 sentences, the French of the 16th century can no longer be considered a V2 language.

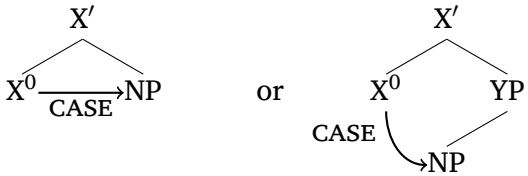
Different explanations for the loss of V2 in French have been suggested in the literature. Adams (1987a,b), for instance, argues that the loss is the result of two overlapping developments. First, subject-initial V2 sentences were reanalysed as SVO sentences. As shown above, the proportion of subject-initial sentences was high in main clauses. Learners may have had this analysis reinforced by the word order in subordinate clauses which featured largely SVO orders (Adams 1987a: 25, 1987b: 86).[41] Due to the high proportion of subject-initial sentences then, learners did not receive sufficient evidence to recognise subject-initial V2 sentences as a derived word order. Instead, learners analysed the sentences in their input as SVO (Adams 1987a,b). The reanalysis, however, was only a necessary development for the loss of V2 according to Adams (1987b: 202).

[41] Note that this claim does not negate the existence of embedded V2 in OF. In fact, Salvesen & Walkden (2017) scrutinised a series of different contexts of complement clauses. Their survey revealed that embedded V2 was possible under certain types of matrix predicates.

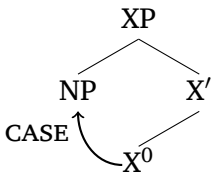
The second relevant change occurred in the prosodic domain. While words and phrases generally bore initial and final stress in OF, initial stress started to weaken in late OF and MF (Adams 1987b: 189). Because the clause-initial constituents in V2 languages are assumed to require stress (Adams 1987b: 134), the new evolving prosodic properties meant that the evidence for V2 was further reduced.[42] That is, the two developments led to a situation in which the acquisition of a V2 was no longer sustainable based on the evidence in the input.

A conceptually related proposal in terms of reanalysis of subject-initial V2 sentences has been developed by Roberts (1993). Framed within the Principles and Parameters framework (Chomsky 1993), Roberts (1993) attributes the loss to a parameter change, specifically in the parameter determining nominative case assignment. The proposal builds on earlier work by Koopman & Sportiche (1991) who argue for two possible configurations under which structural case (which includes nominative) can be assigned; case may either be assigned via government or via agreement. The two options are illustrated in (53). Case assignment under government denotes the configuration where the head assigns case to its complement or to the specifier of the complement (53a).[43] When case is assigned via agreement, the head assigns case to its specifier, as (53b) illustrates.[44]

(53) a. Case assignment under government



b. Case assignment under agreement



(Roberts 1993: 18)

[42] Adams (1987b) also argues that the increasing number of V>2 sentences (XP-S-V) further contributed to the reduction of evidence for V2.

[43] Roberts (1993) defines government as follows:

(i) α governs β iff:
   a. α is a head;
   b. α c-commands β;
   c. there is no head Γ which c-commands β but does not c-command α;
   d. there is no barrier Γ such that Γ includes β but not α.

(Roberts 1993: 19)

For α to c-command β, (i) β must not be dominated by α and (ii) the first X′ that dominates α must also dominate β (Roberts 1993: 19).

[44] Roberts (1993: 19) points out that agreement here is a purely structural notion (spec-head agreement) which does not stand in relation to morphological agreement involving φ-features.

Roberts (1993: 20) argues for three different parameter settings for nominative case-assignment: (i) assignment under agreement only (=(53b)), (ii) assignment under government only (=(53a)) or (iii) assignment under agreement and government (=(53a) & (53b)). Furthermore, the head assigning nominative case to subjects is identified as Agr (or rather the [AGR]-feature borne by it) by Roberts (1993). Setting (iii) enables Agr to undergo movement more readily — or being incorporated in the analysis of Roberts (1993) — as nominative case assignment is not confined to one fixed structural relation as it is the case for (i) and (ii). This will become relevant in Roberts' (1993) account of the loss of V2, as will be shown below.

Similar to other proposals for OF, Roberts (1993) assumes generalised head movement of the finite verb to C. However, because head movement is conceived as incorporation by Roberts (1993), movement of the finite head entails movement of Agr to C as well.[45] Consequently, nominative case is assigned by C in OF. As subjects can either precede or follow the finite verb, the nominative case assignment parameter must be set to assignment under agreement and government — otherwise pre- and postverbal subjects cannot be explained in this model.

The loss of V2 then is the result of a concatenation of different developments. Roberts (1993: 156) proposes that the formation of representations by learners during language acquisition is guided by the so-called *Least Effort Strategy (LES)*. The LES 'instructs' learners to select representations of their input with the shortest chains (i.e. fewest movements) as long as those chains obey the principles of grammar and are consistent with the input. In other words, the LES can be construed as a simplicity bias on a par with similar biases noted for other cognitive domains (cf. Chater & Vitányi 2003). Crucially, the LES played a significant role in the grammatical changes that occurred in French, according to Roberts (1993). First, the LES led learners to reanalyse SpecCP as a position that can function as a target for A-movement in addition to Ā-movement. Subject-initial V2 sentences involve a shorter chain if SpecCP is treated as A-position instead of Ā-position. This is, as Roberts (1993: 157) contents, concomitant with the rise of overt expletives, exemplified in (54).

(54) *Il* **est** judget que nus les ocirum OF
EXPL is judged that we them kill
'It is judged that we will kill them.'
(Roberts 1993: 150)

This change then fed into the next reanalysis that occurred in early MF: Presumably in the 14th century, SVO sentences were reanalysed as AgrP due to the lower number of chain positions this configuration involves (Roberts 1993: 157). Such a reanalysis only became possible because SpecCP had previously assumed its status as A-position. One consequence of this reanalysis is the rise of V3 sentences with initial complements or adjuncts followed by subjects. The existing non-subject-initial V2 clauses, in turn, were

[45] Roberts' (1993: 53) analysis is in fact more complex than alluded to here. V-to-C movement involves a chain of different movements triggered by morphological selection features: C selects Agr, Agr selects T and T in turn selects V, thus resulting in V-to-T-to-Agr-to-C movement.

also reanalysed due to the LES. Instead of movement to SpecCP, initial non-subjects were adjoined to AgrP given the shorter chains this would involve.[46] These successive changes led to an increase of SV sentences. Due to the high frequency of SVO, the evidence for assignment of nominative case under both agreement and government eroded and the parameter was reset to agreement only (Roberts 1993: 153). Crucially, this parameter change is what actually led to the loss of V2 (even though it remained only a minor option beforehand as a result of its decline): C was no longer able to carry [AGR] thereby no longer triggering V-to-C movement, hence all V2 sentences were ruled out by the grammar (Roberts 1993: 198).[47]

A related explanation attributing the changes to a reset parameter can be found in Vance (1997). Following Roberts (1993), Vance (1997) identifies the parameter determining nominative case assignment as the locus of change. What is more, Vance (1997) adopts the same distinction between the decline of V2 and the actual loss of V2. That is, the reanalysis of existing structures as well as the introduction of novel constructions reduced the evidence for V2 (i.e. the decline of V2), eventually resulting in a different parameter setting (i.e. the loss of V2). However, the proposal of Vance (1997) diverges in important aspects from the one in Roberts (1993).[48] Specifically, different factors that drive the change are presumed. Vance (1997) identifies three factors for the reduction of evidence for V2: (i) the high proportion of subject-initial sentences, (ii) the introduction and subsequent increase of V3 sentences and (iii) a reanalysis of non-subject-initial sentences as free inversion. According to Vance (1997), the effect of the preponderance of subject-initial sentences was more indirect as it provided fertile ground for subsequent changes. The second factor, i.e. the rise of V3 sentences, introduced a non-V2 grammar into the population and a grammar competition ensued. Unlike Roberts (1993), Vance (1997) does not consider V3 sentences to be the result of a reanalysis but rather an independent development due to prosodic changes.[49] The increase in V3 sentences meant that CPs were not necessarily projected by speakers as V3 sentences can be interpreted as structures involving IPs. The remaining evidence, which was still sufficient to trigger a V2 grammar in the view of Vance (1997), was further reduced by another reanalysis. Non-subject-initial sentences were analysed as

[46] Roberts (1993) fails to expound how such an analysis became possible in the first place. The LES states that representations need to be consistent with the input. XP-S-V does not fulfil this condition as surface XP-V-S should be maintained. Presumably, the rise of V3 as result of the reanalysis of subject-initial V2 sentences may have contributed to this. However, the LES cannot be adduced as explanation for the observed change.

[47] Roberts (1993: 199) claims that C was still marginally able to select Agr after the parameter was reset, even though confined to a specific set of clause-initial elements. Such a claim is somewhat surprising given that such exceptions would not be expected in this framework.

[48] There are also significant technical differences between the two approaches. The nominative case assignment parameter of Vance (1997) exhibits only two options (instead of three), namely assignment under government (53a) or assignment under agreement (53b). Besides, case assignment may not occur locally in the sense that case can be assigned to chains (Vance 1997: 96). Given that the focus of this thesis does not lie on the development of a technical account of the loss of V2, those differences will not be explored any further here.

[49] Vance (1997) speculates that loss of initial sentence stress might be responsible for the change (cf. Adams 1987b).

cases of free inversion. In contrast to inversion due to V2, free inversion (sometimes also referred to as Romance inversion, Steiner 2014) is characterised by the subject obligatorily following the past participle (Hulk & Pollock 2001: 3). The example in (55) illustrates this property for Modern French. This type of configuration is ungrammatical in the Germanic V2 languages (56a) where the subject has to precede the past participle (56b). Subjects in structures like (55) are typically analysed to remain in the *v*P/VP (Poletto 2016).

(55) Qu'**a** dit *Jean*? FRENCH
what'has said Jean
'What did Jean say?'
(Hulk & Pollock 2001: 3)

(56) a. *Was **hat** gesagt *Johann*? GERMAN
what has said Johann
'What did Johann say?'
b. Was **hat** *Johann* gesagt?
what has Johann said

Consequently, the evidence that a CP is projected was even further reduced. The sketched sequence of reanalyses then created a situation where the evidence for assignment of nominative case under government was diminished to such an extent that the parameter setting was switched. Similar to Roberts (1993), V2 structures were therefore no longer licensed by the grammar, because nominative could no longer be assigned to subjects under government.

Yang (2000, 2002) develops an alternative approach that also relies heavily on learners as the drivers of change. However, instead of positing a reanalysis of V2 sentences, Yang (2000, 2002) argues for a model of grammar competition where learners have two (or more) grammars internalised (see §1.4.1 for more details on the learning model). According to this model, for a community of speakers to lose V2, a non-V2 grammar must have advantage over a V2 grammar in terms of the supporting evidence. Crucially, learners can only rely on unambiguous evidence for the learning task: A V2 grammar is unambiguously identifiable by the presence of XP-V-S-O and O-V-S orders in the input, whereas S-XP-V-O and XP-S-V-O orders constitute clear evidence for a SVO grammar (Yang 2000: 241). The situation in OF was complicated by the existence of *pro*-drop in that non-subject-initial sentences with null subjects were rendered ambiguous — both [XP V *pro*] and [XP *pro* V] were possible structural analyses (Yang 2000: 242). Consequently, learners could only rely on sentences with overt subjects. Yang (2000, 2002) uses counts of SV (including V > 2), VS and *pro* structures by Roberts (1993) to determine whether the V2 grammar or the SVO grammar had advantage over the other. The data reveals that more evidence in favour of a SVO grammar than in favour of a V2 grammar existed (Yang 2000, 2002). Due to this advantage, the V2 grammar was eventually lost in the language community.

A further explanation for the loss of V2 that emphasises learning has been put forward by Steiner (2014). In this analysis, changes in information-structural preferences, in

particular the position of frame-setters within the clause, are the main driver for the loss of V2.[50] As aforementioned, the transition from the 13th to the 14th century saw a significantly increased range of grammatical categories functioning as clause-initial frame-setters. While the clause-initial position in V3 sentences was dominated by adverbial clause frame-setters in the 13th century, other types of frame-setters such as PPs and AdvPs increasingly occupied the clause-initial position in the 14th century. Steiner (2014: 250) analyses these structures as V2 sentences (following Holmberg 2015): Both the finite verb and the constituent preceding the verb were moved to the left periphery (FinP), whereas frame-setters are merged in the left periphery and did therefore not 'count' towards the V2 constraint. A (presumably) independent development leading to a rise of preverbal subjects (cf. Table 1.4)[51] coinciding with this change created an environment in which the evidence for the acquisition of a V2 grammar was significantly weakened (Steiner 2014: 252). The lack of a clear association of frame-setters with a particular grammatical category combined with frequent frame-setter-S-V structures lead to a situation in which V-to-C movement was no longer obvious. Due to pressures of economy, such as the LES of Roberts (1993), learners subsequently assigned simpler representations to the structures in the input: V-to-C movement and an [EPP]-feature borne by Fin (i.e. the defining features of V2) were dropped. According to Steiner (2014: 252), this in turn then can explain why V3 sentences with informational structural categories other than frame-setters in clause-initial position are attested with a high frequency. The [EPP]-feature on FinP acted as a bottleneck as it prevented movement of more than one constituent to the left periphery (cf. §1.2.3.3). In the absence of such a bottleneck, no restrictions applied anymore.

1.3.3 Portuguese

Among the Romance languages, Portuguese has been argued to have retained its V2 grammar for the longest period (Galves 2020: 369). As for other languages, the history of Portuguese is commonly divided into distinct stages (e.g. Ribeiro 1995, Galves 2018). Relevant for tracing the loss of V2 in Portuguese is Old Portuguese (OP; until 15th century),[52] Classical Portuguese (ClP; 16th to 17th century) and early Modern European Portuguese (18th to 19th century). Although the exact point for the loss of V2 is contended (see below), both OP and ClP feature a significant proportion of V2 structures indicative of a V2 grammar (Ribeiro 1995, Cavalcante, Galves & Paixão de Sousa 2015, Galves & Paixão de Sousa 2017, Galves 2020). Based on an examination of 11 texts from the Tycho Brahe Corpus (Galves & Faria 2017), Galves (2020) showed

[50] See Larrivée (2021) for another analysis in terms of information-structural changes.

[51] Steiner (2014) argues that the French of the 14th century was both a V2 language and SVO language. Although this might provide an explanation for the rise of preverbal subjects, it fails to explain why the SVO grammar came into existence in the first place. Besides, the relation between the SVO and V2 grammar remains unclear. That is, are they competing against each other or are both part of the same grammar somehow?

[52] Galves & Kroch (2016: 488) note that Old Portuguese is sometimes further divided into Galego-Portuguese (until end of 14th century) and Middle Portuguese (15th century).

that different types of clause-initial constituents are licensed in ClP such as subjects (57), objects (58) and adverbial phrases (59).[53]

(57) *O duque de Maqueda* **renunciou** o cargo de capitão general CLP
the duke of Maqueda renounced the post of captain general

da armada de Castela
of-the army of Castela
'The duke of Maqueda gave up the post of general of the army of Castela'
(Galves 2020: 375)

(58) Mas povoado **sei** *eu* d'onde elles não haviuam de levar a CLP
but village know I from-where they not had to take the

embaixada debalde.
embassy in vain
'But I know a village from where they would not have to take the embassy away in vain.'
(Galves 2020: 376)

(59) e verdadeiramente *nos* **dá** cuidado o parto da CLP
and truly CL.1.PL gives concern the parturition of-the

Rainha Nossa Senhora
Queen Our Lady
'And we are really worrying about Our Lady the Queen's parturition.'
(Galves 2020: 375)

Interestingly, the amount of clause-initial objects is very low: Only 3% of all non-subject-initial V2 constructions are object-initial (Galves 2020: 377). Despite the low frequency of initial objects, non-subject-initial V2 sentences account for the majority of all V2 sentences. Galves (2020: 377) states that 73% of all V2 sentences are non-subject-initial.[54]

Although the majority of sentences in the ClP texts in the Tycho Brahe Corpus exhibit a V2 order, a non-negligible proportion of deviating structures can be found, as in English and French. On the one hand, sentences with V1 order are attested as the example in (60) illustrates. On the other hand, V3 sentences can be found in the ClP data. Different types of constituents can reside clause-initially. In (61a), a PP and an adverb precede the verb while in (61b) both the object and the subject appear before

[53] This is by no means an exhaustive list. Galves (2020) provides examples for a wide variety of clause-initial constituents.

[54] These numbers should be cautiously interpreted. Galves (2018) analysed sentences with null subjects and a single clause-initial constituent as instances of V2. However, as Yang (2000) notes for French, sentences with null subjects allow for a V2 [XP V NS] and a non-V2 [XP NS V] interpretation (cf. §1.3.2). It is not clear why Galves (2020) opted for the former interpretation. It should also be added that even if all sentences with null-subjects are excluded, the proportion of non-subject-initial sentences remains high (>50%).

Century	V1	V2 (XV)	V2 (SV)	V3
16th	.36	.40	.15	.09
17th	.33	.46	.12	.09
18th	.14	.18	.49	.19
19th	.09	.18	.57	.16

Table 1.5: Proportion of different word orders in Portuguese main clauses by century (Cavalcante, Galves & Paixão de Sousa 2015: 101).

the verb. In the sample of Galves (2020), V1 sentences account for 28% of all structures and V3 sentences for 13% of all sentences.[55]

(60) **Pelejou** *a armada de Holanda* com uma esquadra da CLP
fought the army of Holland with a squadron of-the
armada Real de Castela
army royal of Castile
'The army of Holland fought with a squadron of the royal army of Castile.'
(Galves 2020: 378)

(61) a. E nesta confiança animosamente **soltamos** a vela. CLP
and in-this confidence bravely unloosed.1PL the veil
'And with confidence we bravely unloosed the veils.'
b. e o nome deste soldado também *o tempo* **tem** gastado
and the name of this soldier too the time has wasted
'And time has wasted the name of this soldier too'
(Galves 2020: 381, 382)

Galves (2020) analyses ClP as a relaxed V2 language, following Wolfe (2015b). This means ClP is characterised by verbal movement to the left periphery. The movement of a constituent (or constituents) to a preverbal position is contingent on information-structural aspects (Galves & Paixão de Sousa 2017, Galves & Gibrail 2018, de Andrade & Galves 2019, Galves 2020).[56]

The development of different word orders in ClP and early Modern European Portuguese is scrutinised by Cavalcante, Galves & Paixão de Sousa (2015). Specifically, Cavalcante, Galves & Paixão de Sousa (2015) determined the frequency of V1, V3 as well as subject-initial and non-subject-initial V2 sentences in the Tycho Brahe Corpus, re-

[55] The large proportion of non-V2 sentences has been used as argument against the status of Portuguese as V2 language (Rinke 2009, Sitaridou 2012, Martins 2019). An in-depth discussion of the status of OP and ClP as V2 language is outwith the scope of this thesis.

[56] Galves & Gibrail (2018) also argues that the relative position of postverbal subjects and objects in transitive sentences is sensitive to information-structural concerns.

produced here in Table 1.5.[57] In ClP, non-subject-initial sentences are the most frequent word order, followed by V1 sentences. Interestingly, the frequency of non-subject-initial sentences even increased during the ClP period, as evidenced by the higher proportion in the 17th century. The transition from ClP to early Modern European Portuguese saw significant changes in the frequency of different word order patterns. The previously very frequent V1 and non-subject-initial V2 sentences decrease considerably while subject-initial V2 sentences drastically increased. As a matter of fact, subject-initial V2 sentences even became the most frequent pattern. Crucially, the increase of subject-initial V2 sentences cannot be attributed to a change in the frequency of null subjects. Although the frequency of null subjects fluctuates across centuries, null subjects remain the most frequent type of subject according to data of Galves & Paixão de Sousa (2017: e157).

A technical explanation for the observed changes has been developed by Ribeiro (1995). The explanation is embedded in the Principles and Parameter framework and builds on the assignment of structural case as proposed by Koopman & Sportiche (1991). Structural case may either be assigned under agreement or under government (cf. (53b) and (53a)). Building on the aforementioned proposal by Roberts (1993) for French in §1.3.2 that nominative case assignment is determined by a parameter allowing assignment either under (i) agreement only, (ii) government only or (iii) agreement and government, Ribeiro (1995) argues that nominative case can be assigned either under government or agreement in OP (=(iii)). That both options were possible in OP is illustrated by (62) and (63), respectively.[58]

(62) **Devemos** *nós* a pousar OP
must we to lie down
'We must lie down.'
(Ribeiro 1995: 131)

(63) Entendemos nós [que *a alma* **vive**] OP
understand we that the soul lives
'We understand that the soul lives.'
(Ribeiro 1995: 118)

Akin to the analysis of Roberts (1993) for French, the loss of V2 can then be attributed to a change in the parameter setting of nominative case assignment in Modern Portuguese: Instead of assignment under both government and agreement, nominative can only be assigned via agreement (Ribeiro 1995: 131). To account for the reanalysis, Ribeiro (1995: 131) resorts to Berwick's (1985) Subset Principle according to which learners seek the most restrictive hypothesis. Given that agreement-only is more restrictive than government and agreement, learners changed to this parameter setting. Although this

[57] Note that the values for the proportion of V1 and V3 sentences in the corpus diverges from those reported by Galves (2020). The discrepancy is most likely the result of the smaller sample studied by Cavalcante, Galves & Paixão de Sousa (2015).

[58] Ribeiro (1995: 131) argues only embedded clauses provide unambiguous evidence for assignment of nominative case via agreement since the verb may assign nominative case to the trace of the subject under government in main clauses.

may provide a technical explanation for the loss of V2, it falls short of explaining why other V2 languages have not undergone the same process. A further puzzling aspect is the fact that Ribeiro (1995) argues against an abrupt change following parameter change. To wit, case assignment under agreement became the default with nominative case assignment via government remaining possible as marked option in certain contexts in the 16th century (Ribeiro 1995: 133). This, however, should not be possible if the parameter was previously set to a different value.

A theoretically more neutral explanation for the loss of V2 in Portuguese has been provided by Galves & Paixão de Sousa (2017). Instead of a theory-internal motivation for the change, the loss of V2 is conceived as consequence of changes external to the syntactic realm. An important piece of evidence comes from the placement of clitics attached to the verb in the history of Portuguese. European Portuguese is characterised by a strict enclisis in "nondependent, affirmative tensed clauses where the verb is not preceded by a focalised or interrogative phrase" (Galves & Paixão de Sousa 2017: e168). This can be seen in the contrast in grammaticality between the enclitic and proclitic form in (64a) and (64b), respectively.

(64) a. *O Paulo* **falou**=me. EUROPEAN PORTUGUESE
the Paul speak.3SG.PST=1SG.DAT.CL
'Paul spoke to me.'
b. **O Paulo* me **falou**
the Paulo 1SG.DAT.CL speak.3SG.PST
(Galves & Paixão de Sousa 2017: e168)

In ClP in contrast, the position of clitics was not categorically enclitic.[59] In fact, proclisis was the dominating pattern prior to 1700 (Galves & Paixão de Sousa 2017: e168), exemplified in (65) by the object clitic *o* 'it' preceding the verb. Galves & Paixão de Sousa (2017: e170) propose that the variable order of clitics in ClP is an instantiation of the Tobler-Mussafia law. That is, clitics were barred from appearing in initial position in an intonational phrase. The information-structural status of the clause-initial constituent therefore effectively determined whether proclitics or enclitics were used: Proclitic forms were used with different types of topics and focused constituents, whereas enclitic forms were used when the clause-initial constituent expressed contrast (Galves & Paixão de Sousa 2017: e169, cf. Galves, Brito & Paixão de Sousa 2005). Since contrastive elements form their own intonational phrase (cf. Frascarelli & Hinterhölzl 2007), proclitics would have been the initial element in the intonational phrase with the verb, consequently violating the Tobler-Mussafia law.

(65) *O Evangelho* o **diz**: Erunt signa in sole, et luna, … CLP
the Gospel 3SG.ACC.CL say.3SG.PRS
'The Gospel says it: Erunt signa in sole, et luna, …'
(Galves & Paixão de Sousa 2017: e169)

[59] This only applies to V>1 sentences. Galves & Paixão de Sousa (2017: e168 fn.20) observe strict enclisis in V1 clauses during all stages of European Portuguese.

According to Galves & Paixão de Sousa (2017), the change from a variable, phonologically-conditioned system to a fixed system was a consequence of phonological change affecting pretonic vowels. As result of this change, secondary stress aligned with the first syllable of the word. Considering the inherently unstressed nature of clitics, proclisis became unsuitable and enclisis increased in frequency (Galves & Paixão de Sousa 2017: e172).[60]

Relevant for explaining the loss of V2 is the fact that the increase of enclisis coincided with the decline of VS structures. Galves & Paixão de Sousa (2017) conjecture a connection between the former and the latter: Due to the increase of enclitic forms, the originally marked status of XPV-cl and especially SV-cl was lost. Markedness of syntactic structures is tied to their prosodic contours: if a contour accounts for the majority of all patterns, it is unmarked (Hinterhölzl 2009: 51). Hence, the increase of enclisis rendered SV-cl unmarked. Galves & Paixão de Sousa (2017: e172) assume that the intonational contour of the clause-initial constituent was no longer independent from the following constituents as a consequence of the shift in markedness. This in turn led to a reanalysis of the position of preverbal subjects.[61] Instead of targeting a position in the left periphery, subjects were realised in what Galves & Paixão de Sousa (2017: e172) refer to as 'subject position'. Albeit not further specified, this position must be below the CP-domain, presumably SpecTP.[62] According to Galves & Paixão de Sousa (2017: e173), the reanalysis of the subject position subsequently led to the loss of V-to-C movement. The last change effectively meant that Portuguese was no longer a V2 language (cf. §1.2.3.3). Before turning to the situation in Welsh, it is worth pointing out that the account by Galves & Paixão de Sousa (2017) dates the loss of V2 later than the one by Ribeiro (1995). Given the prevalence of XVS structures before the 18th century (Table 1.5) however, a later date, as suggested by Galves & Paixão de Sousa (2017), might be more appropriate.

1.3.4 Welsh

The final language whose loss of a V2 grammar will be considered here is Welsh. While the unmarked word order in Modern Welsh is VSO — akin to other Celtic languages (with the exception of Breton, cf. §1.2.2) — earlier stages of Welsh, specifically Middle Welsh (12th to 15th century), have been described as V2 language (Willis 1998, 2007, Meelen 2016). In so-called *abnormal* sentences, which correspond to affirmative declarative sentences, the verb typically forms the second constituent of the clause.[63,64] As in the other V2 languages, a wide range of different constituents can occur in the preverbal

[60] Given the incomaptibility of clitics and stress, one might expect that such a change would entail a swift switch from proclitics to enclitics. The data from Galves & Paixão de Sousa (2017: e170) suggests though that no abrupt changes occurred. It was only in the 19th century that enclisis became fully categorical in European Portuguese. This issue must be left open for future research to address.

[61] An important question that is left unanswered is the decline of non-subject-initial sentences.

[62] See de Andrade & Galves (2019) for an account of the changes in the articulated left periphery.

[63] According to Willis (1998: 4), the term *abnormal* is owed to the archaic status of abnormal sentences in Modern Welsh.

[64] See Meelen (2016) for an overview of possible word orders in affirmative declarative main clauses other than abnormal sentences.

position such as subjects (66a), objects (66b) and adverbs (66c), among others (Willis 1998: 51–52). Apart from a small number of exceptions, finite verbs are preceded by the particles *a* or *y(d)*. The choice of the particles is determined by the type of the clause-initial constituent: *y(d)* mostly co-occurs with clause-initial adjuncts (including subordinate clauses), whereas *a* is selected with all other clause-initial constituents (Willis 1998: 52, 2007: 436–437, Meelen 2016: 115–116).

(66) a. A' *r ederyn* a **doeth** y 'r ynys honn. MIDDLE WELSH
and the bird PRT came to the island this
'And the bird came to this island.'
b. Ac ystryw a **wnaeth** *y Gwydyl*.
and trick PRT made the Irish
'And the Irish played a trick'
c. Yn Hardlech y **bydwch** seith mlyned ar ginyaw …
in Harlech PRT be.FUT.2PL seven years at dinner …
'In Harlech you will be at dinner for seven years …'
(Willis 1998: 51)

The word order in Middle Welsh was not strictly V2 in abnormal sentences as exceptions have been observed. Adverbs could be placed either before or after preverbal subjects and objects (Willis 1998, Meelen 2016: 118).

A further type of affirmative declarative main clause with V2 word order are *mixed* sentences. This type is superficially similar to abnormal sentence, as illustrated in (67): The verb is preceded by the same particles as in abnormal sentences (i.e. *a* and *y(d)*) which are determined by the type of clause-initial constituent (Willis 1998: 4). From a information-structural perspective, however, the two sentence types have traditionally been argued to differ. The clause-initial constituent in mixed-sentences is focused, whereas in abnormal sentences the clause-initial constituents constitute topics (Meelen 2016: 285).[65] Besides, the finite verb does generally not agree with the subject and exhibits a default third-person singular inflection instead (Willis 1998: 5, Meelen 2016: 119).

(67) a. bydhawt ragot ti gyntaf yd **agorawr** y porth MIDDLE WELSH
be.FUT.3SG to.2SG you first PRT open.IMPER the gate
'For you shall the gate be opened first.'
b. Oed maelgun a **uelun** in imuan
be.PST.3SG Maelgwn PRT see.PST.1SG PROGR fight.INF
'It was Maelgwn that I could see fighting.'
(Meelen 2016: 119)

The distribution of different constituent types in clause-initial position as well as their historical development in Middle Welsh has been studied in more detail by both Willis

[65] Meelen (2016) discovered that the distinction is not as clear-cut as previously assumed. The distinction of abnormal and mixed sentences lies outside the scope of this thesis and the interested reader is referred to Chapter 6 in Meelen (2016).

(1998) and Meelen (2016). The observed distributions resemble each other and both scholars arrive at concurring conclusions: First, the proportion of clause-initial objects is already low in Middle Welsh and declines even further during that period. Second, verbal nouns occur with considerable frequency clause-initially, 20% on average according to Meelen (2016: 306). Akin to objects however, the frequency of clause-initial verbal nouns decreases in later Middle Welsh sources. Third, adjuncts constitute the most frequent non-subject-initial constituent type. Interestingly, in some of the examined texts adjuncts even make up a higher proportion of clause-initial constituents than subjects. Towards Early Modern Welsh, adjuncts in clause-initial position are in decline though. Finally, subjects account in most cases for the highest proportion of clause-initial constituents. Contrary to other types occurring clause-initially, the proportion of subjects actually increases towards the end of Middle Welsh. That is, towards the end of the Middle Welsh period, V2 sentences were predominantly subject-initial.

The shift in the composition of the clause-initial position from the transition from Middle Welsh to early Modern Welsh was significant but the loss of V2 was potentially induced by a further development. In the 16th century, the preverbal particles *a* and *y(d)* started to be omitted. Albeit originally a purely phonological phenomenon, the omission of the particles became increasingly widespread in the latter half of the 16th century (Willis 1998: 188–189). This development had far-reaching consequences for the V2 grammar in Welsh. In the account of Willis (1998), the clause-initial constituent and the preverbal particles are in an agreement relationship. That is, the particle (with adjoined verb) and the preceding constituent must be in a Spec-head relation. Willis (1998: 183) argues that this agreement relation provides crucial evidence to learners for movement of the clause-initial constituent to a specifier. Learners can further deduce that this position must be SpecCP, because the agreeing particles are complementisers. Moreover, the identical preverbal particle for objects and subjects signals learners that subjects and objects must reside in the same structural position thereby further strengthening the evidence for movement to CP (Willis 1998: 183). In the light of the important role these particles played during language acquisition, their loss caused a significant reduction of evidence for the V2 grammar. According to Willis (1998: 193), the amount of evidence for learners became in fact so low that the acquisition of the V2 grammar was no longer sustainable. The ensuing reanalysis of the input by learners was driven by the Least Effort Strategy of Roberts (1993): Adjuncts no longer occupied SpecCP and were adjoined instead (Willis 1998: 190). Similarly for subject-initial sentences, no CP had to be projected anymore as structures could be analysed involving the IP-domain only (Willis 1998: 192). Following these reanalyses, Welsh could no longer be considered a V2 language.[66] In what follows, I will synthesise the different approaches to the loss of V2 in English, French, Portuguese and Welsh.

1.3.5 Parallels in the loss of V2

In the past four subsections, the loss of V2 in English, French, Portuguese and Welsh was examined. The review of different approaches to the loss of V2 has shown that not

[66] See Willis (2007) for an account of the developments after V2 was lost.

much consensus has been reached in the literature for individual languages, let alone cross-linguistically. Different causes have been suggested, ranging from phonological changes, morphological changes and changes to the information-structural organisation of languages to economy principles disfavouring specific aspects of a V2 grammar and language contact. Nonetheless, some aspects recur in different analyses such as the Least Effort Strategy of Roberts (1993). Furthermore, common themes have crystallised. First, in none of the sketched analyses V2 was lost in and of itself, because it was too hard to learn. Although some stages in the process of losing V2 have been argued to lack sufficiently clear patterns for learners (cf. van Kemenade & Westergaard 2012), none of the 'healthy' initial stages are considered too difficult to maintain. Even Ribeiro (1995), who argues for a reanalysis on grounds of simpler available parameter settings, refers to a specific aspect of the V2 grammar (i.e. nominative case assignment) and not the whole grammar. Besides, the continuity and strengthening of V2 systems in the Germanic languages apart from English does contradict the hypothesis that V2 is too hard to learn. The second observation that can be made is that the driving factors for the loss of V2 appear to be extraneous in the sense that they are not directly connected to the syntactic domain itself. This is best illustrated by the proposed phonological and morphological factors responsible for the loss. Changes in these domains interface with syntax in such a way that the V2 system becomes affected. In particular, the evidence for V2 is argued to deteriorate, preventing the acquisition of the V2 grammar. This, in turn, is connected to the third observation. Language learning and the corresponding evidence play a crucial role in the loss of V2. This becomes evident by the frequent invocation of *reanalysis* and/or learner-internal grammar competition to explain the loss. Interestingly, all of the proposed mechanisms of change are unified in that they seem to cause a redistribution of clause-initial constituents. That is, although the assumed processes are different, an increase in subject-initial V2 sentences appears to be the result. This stands in contrast to earlier stages of the languages where the clause-initial position was characterised by a more diverse distribution (i.e. more non-subjects). The striking parallel across all languages that have lost V2 begs the question whether these shifting distributions might be the proximal cause of the loss of V2. That is, under such a view, the shifting distributions would be the immediate (i.e. proximal) cause, while the factors causing the redistributions would be the distal causes for the loss. If this was indeed the case, I would be in a position to develop an analysis of the loss of V2 that can explain the developments in all these languages. Note that many of the accounts also highlight the rise of V3 constructions. In the remainder of the present thesis however, I will focus solely on the distributional change in the clause-initial position as explanation. Past research has provided evidence for domain-general benefits of variability in the input for learning (Raviv, Lupyan & Green 2022). Applied to the question at hand, these findings suggest that low variability in the clause-initial position should hinder the acquisition of V2, hence leading to the loss of V2.[67] According to this interpretation then, the rise of V3 does not exhibit the same status as the distribution of the clause-initial constituents.

[67] This hypothesis will be discussed in significantly more detail in §1.5.

However, before such an analysis can be explored in more detail, one fundamental question has to be addressed: Although the diachronic developments suggest that changes elsewhere in the grammar caused a shift in the distribution of clause-initial elements, it is far from clear whether the rise of subject-initial sentences is actually the cause or the effect of the loss of V2. In other words, the rise might either be the driver of the loss of V2 or the result of the loss of V2. Crucially, this only refers to the proximal cause of the loss of V2. The presumed distal causes for the loss of V2 — i.e. the language changes that may have triggered the distributional changes in the first place — are not considered here. As aforementioned, the goal is to provide an analysis that can explain the developments across languages. This would not be possible when the different distal causes are examined.[68] Returning to the question of cause and effect, I am effectively dealing with the infamous chicken-or-egg problem with no realistic way of distinguishing the two options based on the data alone. One avenue worth exploring in order to solve this problem is the connection between subject-initial sentences and learning. If subject-initial sentences impact learning, the effect should be predicted by models that attribute language change to language acquisition.

To summarise, many causes for the loss of V2 have been proposed. The drivers of the change are language-specific and the loss of V2 appears to be an epiphenomenon of changes that have taken place somewhere else in the grammar. Furthermore, a similar pattern has emerged across the four languages surveyed in this section in that the proportion of subject-initial sentences significantly grew over time. This is also relevant against the backdrop of analyses that have invoked learning as a relevant factor. The shifting distributions in combination with learning could thus be used to develop a cross-linguistic explanation for the loss of V2. However, the high amount of subject-initial sentences constitutes a chicken-or-egg problem in the context of the loss of V2. Models of language change incorporating learning should make predictions that help identifying cause and effect. In the next section, I will therefore highlight three models of learning where the nature of the input is connected to language change.

1.4 The role of learning in the loss of V2

Different causes for the loss of V2 have been identified in the literature, although the factors responsible for it appear to be language-specific. There is, however, one feature that characterises the trajectory of all diachronic developments with regards to the loss of V2: the rise of subject-initial clauses. Moreover, many accounts allude to some form of learning processes that play a role in the change as well. Two questions can be raised in response to these observations: First, what actually is the contribution of learning? And second, what role does the increasing frequency of subject-initial clauses play? The shifting distribution may either be the cause or the effect of change (i.e. a chicken-or-egg-problem). This section will introduce three models that incorporate both learning and the nature of the input, namely the variational learning model (§1.4.1),

[68] Future work could investigate whether the proposed mechanisms actually lead to a rise in subject-initial sentences. Such studies, however, would differ from the current one. Besides, they might not necessarily implicate learning but language usage.

cue-based learning (§1.4.2) and the micro-cue model (§1.4.3).[69] According to these models, the high proportion of subject-initial clauses are the cause and not the effect of the loss of V2. Hence, the models provide a general explanation that holds for all observed changes. Note that the goal of this section is not to trace and account for changes in individual languages. Rather, the general interaction of input and learning will be highlighted.

1.4.1 Variational learning model

One approach that underscores the role of learning in interaction with the input is the variational learning model (Yang 2000, 2002, 2010). The inspiration for the model derives from the process "of natural selection in biological systems" (Yang 2010: 1162). Instead of selecting between genotypes however, the variational learning model assumes competition between different grammatical hypotheses (Yang 2000, 2002, 2010).[70] The hypothesis space is provided by Universal Grammar (UG). That is, learners are tasked with converging on appropriate hypotheses by selecting from the provided set. Relevant for the goals of this section is the assumption that learners will rely on their input to accomplish this task (Yang 2000, 2002): Each hypothesis (or grammar) is associated with a weight (i.e. probability). When encountering a sentence in the input, the child will perform a random weighted pick of a grammar G_i from the hypothesis space that will then be used to parse the input structure. If the grammar G_i can parse the sentence, G_i will be rewarded in that its weight will be increased. At the same time, other grammars will indirectly be punished. If, however, the input string cannot be parsed by G_i, the grammar will be punished and all other grammars indirectly rewarded. This process is repeated every time learners receive input.

If learners were exposed to an idealised homogeneous linguistic environment, the target grammar should never be penalised and learners would pick the correct grammar from the hypothesis space. If, however, the linguistic environment is more heterogeneous, a different picture will emerge. According to Yang (2000: 237), different events may introduce new varieties into the linguistic environment of learners thereby diversifying the input of learners. For instance, L2 speakers might be introduced into the linguistic community or an innovation might spread from a social niche to the broader community. As a result, two grammars G_1 and G_2 will be present in the input. Applying this to the phenomenon under discussion, G_1 could represent a V2 grammar and G_2 a SVO grammar. Learners will proceed as aforementioned during language acquisition: A grammar is randomly selected based on its weight for parsing a structure in the input. It will be rewarded if it can parse the input, otherwise the grammar will be punished. In

[69] See also Cournane & Klævik-Pettersen (2023) for a further alternative that appeals to learning and the evidence in the input to explain the loss of V2. Unfortunately, their analysis was published too close to completion of this thesis to be discussed it in more detail.

[70] The nature of these grammatical hypotheses differs across publications. Whilst Yang (2000, 2002) argues for these hypotheses to be grammars, Yang (2010) posits that the hypotheses are on a smaller scale, namely parameters. The exact nature of these grammatical hypotheses does not bear relevance for the further discussion. I will therefore refrain from committing to one of the two notions here though I will adopt the terminology of Yang (2000, 2002).

certain cases, input structures will be ambiguous and can thus be parsed by both grammars. In the context of G_1 and G_2, these structures might correspond to SVO clauses.[71] Whichever grammar is selected, it will be rewarded. Importantly, extended exposure to ambiguous structures will provide neither grammar G_1 nor grammar G_2 with an advantage — none of the grammars can effectively gain the upper hand. In this light, the observed rise of subject-initial sentences entailed that a very large proportion of learners' input became unsuitable for distinguishing a V2 grammar from the competing SVO grammar.

What matters for distinguishing different grammars are thus structures that can only be analysed by one grammar. The proportion of sentences that can only be analysed by grammar G_1 is, what Yang (2000, 2002) refers to as advantage α of G_1 over G_2. Likewise, the proportion of sentences that can only be parsed by grammar G_2 is the advantage β of G_2 over G_1. Yang (2000, 2002) proves mathematically that if $\beta > \alpha$, G_2 will eventually take over.[72] Over the course of multiple generations then, the advantage of the SVO grammar will constantly increase (Yang 2000: 239). This rests on the assumption that learners produce structures of a grammar with the same weight as they select the grammars for parsing (Yang 2002: 27). Consequently, the advantage of the SVO grammar in production will also be larger than the one of the V2 grammar. Returning to the rise of subject-initial sentences, the declining variation in the clause-initial constituent reduced the amount of evidence for V2. That is, V2 was further disadvantaged by the increasing number of subject-initial sentences. Given that structures with non-V2 order were also on the rise in many languages, the advantage for the SVO grammar exceeded the one of the V2 grammar. Taken together, the rise of subject-initial sentences is the cause, not the effect in the variational learning model as they rendered a significant proportion of the input ambiguous. Combined with a concurring increase in V3 structures compatible only with a SVO grammar, the unambiguous evidence for a V2 grammar was reduced to such an extent that it no longer had an advantage over the SVO grammar.[73]

1.4.2 Cue-based learning

Cue-based learning, developed in Lightfoot (1999, 2006), constitutes another approach underscoring the role of the input for language change. This approach centres around the dichotomy of I(nternal)-language and E(xternal)-language (Chomsky 1986): While

[71] Yang (2002: 35) argues that no sentence type could unambiguously identify a V2 grammar.

[72] In a hypothetical scenario where a SVO and V2 grammar compete, the presence of a single XPSVO sentence among exclusively SVO sentences would in theory suffice to give the SVO grammar an advantage over the V2 grammar. Although this completely hypothetical example is very unlikely to ever occur, one may ask whether a model where the repercussions of parsing a single sentence could be so significant is desirable at all. Lightfoot (2006: 82) points out that children are very unlikely to deviate from the language of the previous generation due to a single ill-formed utterance (cf. also §1.4.2). This, of course, does not mean that the variational learning model should be discarded. Rather some measure of certainty should be added to the model. Besides, Yang (2002: 31) emphasises that learners do not actually track frequencies during learning.

[73] The variational learning model was discussed in the context of V2, however it has also been applied to other situations of language change. For instance, Heycock & Wallenberg (2013) apply the model to data from Icelandic and Swedish to test the loss of V-to-T movement.

I-language captures the notion of *grammar*, that is the mental representation(s) of a speaker's (or signer's, for that matter) linguistic knowledge (Lightfoot 2006: 7), E-language is conceived as the result of using I-languages (Lightfoot 2006: 12). For instance, notions like *German* or *English* are what Lightfoot (2006: 13) considers E-language. Under this view, a speaker can change their usage of their grammar but crucially not the grammar itself (Lightfoot 1999: 84). Changes to the E-language, however, can have what Lightfoot (1999: 89) has labelled 'catastrophic' effects for the next generation of speakers: E-language forms the primary linguistic data (PLD) of learners. If changed usage entails that a threshold for the emergence of a new grammatical property is crossed, the I-language (i.e. grammar) of a learner deviates from that of other speakers in an abrupt fashion and hence constitutes a catastrophe (Lightfoot 1999: 91). Although changes in the grammar occur abruptly, they may not spread immediately in the population and the diffusion of the change may proceed gradually (Lightfoot 1999: 104).

To account for these ideas, Lightfoot (1999, 2006) argues that children rely on cues during learning. This means that children search the mental representations they have created by parsing the input, for specific structures, so called cues.[74] The cues themselves are provided by UG and do not constitute full structures but merely pieces of structures (Lightfoot 2006: 78).[75] Parsing a sentence like (68), for instance, will yield a representation for the word order in the VP along the lines of $_{\text{VP}}$[V DP]. Hence, (68) expresses the cue for head-directionality in the VP (Lightfoot 2006: 78).

(68) The mouse gobbled the cheese.

Crucially, a cue is only expressed in a sentence "if the cue is unambiguously required for the analysis of the sentence" (Lightfoot 2006: 78). Different languages may express the same cue differently (Lightfoot 2006: 78). For example, a cue *C* in language A may be expressed by declarative sentences, yet language B expresses *C* unambiguously only in exclamative sentences. Lightfoot (2006: 79) also notes that the representations formed by learners may not necessarily be adult-like in that the input is only partially analysed. However, as the acquisition process progresses, the representations become increasingly more abstract (Lightfoot 2006: 80). This also entails that the learning process must follow a clear trajectory (Lightfoot 1999: 150, 2006: 79). Returning to the example in (68), the learner needs to establish the phrasal categories first before they can determine whether the language has OV or VO order. That is, the learner has to identify *gobble* as verb and *the cheese* as DP before analysing (68) as an instantiation of VO.

As mentioned above, grammar 'catastrophes' are argued to occur when a certain threshold for a novel grammatical structure is met. In the context of cue-based learning, this means that the cues need to be sufficiently expressed in the E-language. Lightfoot (2006: 82) motivates this by learners' general resiliency against ungrammatical structures: Children do not converge on a novel grammar after being exposed to a single

[74] Lightfoot (1999, 2006) further argues that learners follow degree-0 learnability (cf. Lightfoot 1994). That is, learners can only consider unembedded clauses as sources for language acquisition.

[75] This accords with the approach of Fodor (1998) where so-called *treelets* (i.e. small pieces of tree structures) are considered to be triggers of grammatical structures.

ungrammatical utterance. Lightfoot (2006: 82) adduces the example of a L2 English speaker with German as L1 who produces an embedded sentences with OV order instead of VO in the presence of a learner. The learner will not learn an OV grammar after a single exposure.[76] This then raises the question when exactly a cue is sufficiently attested. Lightfoot (1999: 154) rejects the idea that uniform thresholds can be established and argues that thresholds need to be established on a case-by-case basis instead.

One of the constructions discussed in more detail by Lightfoot (1999) is V2. According to Lightfoot (2006: 86), the cue for V2 can be any phrasal category in the CP followed by the finite verb, i.e. $_{CP}$[XP $_{C}$V]. Lightfoot (1999: 154) delimits the availability of the V2-cue to non-subject-initial sentences though, as the cue is only unambiguously expressed in those contexts given that subject-initial sentences permit different analyses (SVO or V2, cf. Yang 2000, 2002). Based on counts from Lightfoot (1995) and Ans van Kemenade (p.c.), Lightfoot (1999: 156) gauges the threshold for the acquisition of V2 between 17% and 30%. This means that learners' input has to feature 17%–30% of non-subject-initial sentences for the V2-cue to be expressed. This threshold can immediately explain how an increasing amount of subject-initial sentences affects the acquisition of a V2 grammar: Learners do not find sufficient evidence for the V2-cue ($_{CP}$[XP $_{C}$V]) in the input. This lack of evidence then prevents learners from acquiring a V2 grammar. That is, similar to the variational learning model, the rise of subject-initial sentences is the cause and not the effect of the loss of V2. In the next subsection, an approach related to the cue-based learning will be examined.

1.4.3 Micro-cue model

The cue-based learning approach of Lightfoot (1999, 2006) is extended by Westergaard (2008, 2009b, 2014). The new model, referred to as micro-cue model, abandons the notion of major or large cues (or parameters) in favour of smaller cues. This is motivated by the general failure of parametric approaches to capture language-internal variation. In the context of V2, for instance, setting the parameter to [+V2] would predict that all main clauses (at least) should follow a V2 word order, whereas the opposite setting (i.e. [-V2]) should block any V2 structures in a language. English, as aforementioned, has lost its V2 property during the Middle English period. Hence, present-day learners of English would set the V2 parameter to [-V2]. Lightfoot & Westergaard (2007) point out, however, that structures like the ones given in (69) are problematic for parametric approaches such as Lightfoot's cue-based learning. All examples in (69) exhibit the characteristics of a V2 order with verbs occupying the second position and non-subjects the clause-initial position.

(69) a. Which cheese **did** *the mouse* like?
b. None of the cheeses you normally eat **would** *a mouse* like.
c. At the cheese monger's **was** *the mouse* always welcome.

[76] Lightfoot (2006: 82), however, somewhat qualifies this statement a few sentences later in that they argue that one sentence expressing a cue might suffice in certain circumstances.

(70) a. *Which cheese *the mouse* **did** like?
b. *None of the cheeses you normally eat *a mouse* **would** like.
c. At the cheese monger's, *the mouse* **was** always welcome.

Owing to the fact that V2 is obligatory in English *wh*-questions (69a) and negated phrases (69b) — evidenced by their ungrammatical V3 equivalents in (70) — Rizzi (1996) coined the term residual V2 language.[77] Although locative inversion structures such as the one in (69c) do not exhibit the same obligatory nature as residual V2 structures (70c), they are nonetheless unpredicted under a general V2 parameter or cue.

Another piece of evidence extensively discussed by Westergaard (2008, 2009b) pertains to variation in North Norwegian, specifically in the dialect of Tromsø. While Standard Norwegian is characterised by a strict V2 order in *wh*-questions, the dialect of Tromsø (akin to other Norwegian dialects) allows both V2 and non-V2 orders (Westergaard 2003, Westergaard & Vangsnes 2005, Westergaard 2009a): This is argued to be a case of conditioned variation where the length of the *wh*-element as well as the information-structural status of the subject determines the word order: While *wh*-questions with disyllabic or long *wh*-elements show obligatory V2 order (71), questions with monosyllabic *wh*-elements (*ka* 'what', *kem* 'who' and *kor* 'where') display non-V2 when the subject is informationally given (72) and V2 when the subject is informationally new (73) (Westergaard 2003, Westergaard & Vangsnes 2005).[78]

(71) a. Ka slags rødvin **vil** *du* ha? NORWEGIAN (STANDARD)
which kind red.wine will you have
'What kind of red wine would you like?
b. *Ka slags rødvin *du* **vil** ha?
(Westergaard 2009a: 51)

(72) Kor *vi* **lande** henne? NORWEGIAN (TROMSØ)
where we land LOC
'Where should we land?'
(Westergaard 2009b: 25)

(73) Kor **er** *mitt fly*? NORWEGIAN (TROMSØ)
where is my plane
'Where is my plane?'
(Westergaard 2009b: 25)

Akin to English, a general V2-cue cannot account for the attested pattern in Norwegian. However, unlike English, one could use a grammar competition account such that a V2 grammar and a non-V2 grammar compete in the dialect of Tromsø. Westergaard (2008:

[77] The term *residual V2* of course also incorporates the fact that historic stages of English were characterised by a V2 grammar (cf. §1.3.1). In fact, multiple languages have been described as residual V2 language for similar reasons, e.g. French and Spanish (Rizzi 1996, Holmberg 2015: 344). However, see Cruschina & Sailor (2022) for arguments against conflating formal and historical senses of residual V2.

[78] Note though that Westergaard (2009a: 56) also observed considerable variation between speakers, suggesting the influence of sociolinguistic variables.

1853) adduces evidence from child-directed speech and L1 acquisition against such an account. The amount of evidence for V2 across contexts (54.2%) exceeds the evidence for non-V2 significantly (9.6%) in child-directed speech in Tromsø (Westergaard 2008: 1852). The cue-based learning approach as outlined by Lightfoot (1999) would predict that V2 orders are learned much earlier than the non-V2 ones. Westergaard (2008, 2009b) demonstrates, however, that this prediction is not borne out as children show target-consistent behaviour from early on picking up fine distinctions. Therefore, she argues that cue-based learning is inadequate for explaining the observed variation both in Norwegian and English.[79,80]

The micro-cue model is equipped to deal with such language-internal variation. By departing from major cues such as a general V2-cue ($_{CP}$[XP $_{C}$V]) in favour of smaller micro-cues, the model can handle variation without seeking recourse in a grammar competition model. According to the micro-cue model, learners parse the input and form small syntactic structures, the micro-cues (Westergaard 2014). These structures are the result of an interaction of the input, UG and third factors (Chomsky 2005), such as principles of economy (Westergaard 2009b: 64, Westergaard 2014: 38). In contrast to Lightfoot's (1999, 2006) model, UG does not provide the cues but rather the necessary principles of structure building and syntactic primitives such as features and categories (Westergaard 2009b: 52, 2014: 38). The formed micro-cues are then incorporated into the grammar, though the actual procedure remains vague (Westergaard 2009b: 65). That is, micro-cues become "part of a speaker's internalized knowledge of a specific language" (Westergaard 2017: 460).[81]

Given the sensitivity to the linguistic context displayed by learners, micro-cues must encode the contexts in which they occur (Westergaard 2008, 2009b, 2014). In the context of V2, this means that V2 is conceived as conspiracy — that is, V2 does not constitute a unified phenomenon but rather consists of multiple, domain-specific rules (Lohndal, Westergaard & Vangsnes 2020, cf. also Weerman 1989). Following Westergaard & Vangsnes (2005), Westergaard (2008, 2009b) proposes a split-CP approach (cf. §1.2.3.3) in which ForceP is expressed differently, depending on the clause type (e.g. DeclP in declaratives and IntP in *wh*-questions). This then provides the context for the micro-cues. In the dialect of Tromsø, parts of the V2-grammar can then be accounted for by micro-cues for declarative main clauses and long *wh*-questions, provided in (74a) and (74b), respectively.

(74) a. $_{DeclP}$[XP $_{Decl°}$V]
b. $_{IntP}$[XP$_{[+wh]}$ $_{Int°}$V]
(Westergaard 2009b: 60)

[79]The early target-like productions of children are not unique to V2 but have been observed for clause-internal subject positions and the position of possessors in DPs too (Westergaard 2011, Anderssen & Westergaard 2010).

[80]Westergaard (2017: 458) also calls into question that significant E-language changes can occur without previous changes in the I-language.

[81]Compared to the proposals by Yang (2000, 2002) and Lightfoot (1999, 2006), the micro-cue model has the advantage that not all cues or grammars/parameters must be provided by UG (Westergaard 2014: 33). This allows for a slimmer language faculty — an aspect that is preferable under Occam's razor.

Micro-cues can also encompass specific lexical items. Some word order variation in Norwegian is contingent on lexical items. The word order in sentences with clause-initial *kanskje* 'maybe' can either be V2 or non-V2 (75) (Westergaard 2008, 2009b).[82] The corresponding micro-cue can thus be formulated as in (76).

(75) Kanskje (**kommer**) *kongen* (**kommer**). NORWEGIAN
maybe come.PRS king.DEF come.PRS
'Maybe the king is coming.'
(Westergaard 2009b: 20)

(76) $_{\text{DeclP}}$[*kanskje* XP …$_{\text{VP}}$[V]]
(Westergaard 2009b: 60)

One immediate consequence of the micro-cue model is the way input frequencies need to be calculated. Instead of subsuming all V2 cases, frequencies need to be considered for each cue individually (Westergaard 2008: 1857, 2009b: 208). The effect is non-trivial. For instance, the frequency of sentences with clause-initial *kanskje* 'maybe' such as (75) changes drastically from 1.9% to 12% when only non-subject-initial sentences are considered instead of the whole set of sentences in the input (Westergaard 2009b: 209). Despite the context-sensitivity, learners also need to generalise to some extent as grammars are productive and not merely the accumulation of specific constructions (Westergaard 2014: 40). In the micro-cue model, generalisations proceed in small steps by extending existing cues. For instance, learners might add a new feature or (sub)category to a micro-cue (Westergaard 2014: 40).[83]

Similar to the cues in the model of Lightfoot (1999, 2006), micro-cues are only expressed in unambiguous structures (Westergaard 2014: 37). For instance, learners will find the micro-cue for declarative V2 sentences only in non-subject-initial clauses.[84] Unlike cue-based learning however, micro-cues do not need to be attested with the same degree of robustness in the input. Westergaard (2011) shows that learners are able to acquire a fine distinction between postverbal pronominal and nominal subjects relative to sentence adverbs and negation despite the low frequency.[85] Frequency does still play an important role though: Low frequencies of micro-cues enable language change because constructions become susceptible to change (Westergaard 2009a, 2011, 2021a)

[82] This can be seen as further argument against a parametric approach. It seems highly unlikely that UG provides cues that are specific to certain lexical items (Westergaard 2008: 1859).

[83] The micro-cue model is thus situated between traditional parametric approaches and constructionist theories (Westergaard 2009b, 2014, cf. also Westergaard 2009c). Both the micro-cue model and constructionist approaches assume a development from specific to general, but no genetic endowment is stipulated by the latter models (Westergaard 2014: 38).

[84] According to Westergaard (2009b), subject-initial V2 clauses are not simply discarded since they can convey other useful information to learners in certain cases. If those clauses contain negation or clause-medial adverbs, subject-initial sentences will provide evidence for V-to-I movement at least.

[85] One could simply dismiss this finding by assuming that this particular phenomenon does require a lower frequency than V2 for instance. This would be compatible with what Lightfoot (1999) has claimed. However, such an account is unsatisfactory since it falls short of predicting any frequency thresholds.

— children may ignore a construction due its low frequency for example (Westergaard 2011: 323).

Against the backdrop of the micro-cue model, the role of subject-initial sentences is not as straightforwardly accountable as in the other two models discussed above. The decline of non-subject-initial clauses concomitant with the rise of subject-initial V2 clauses should not immediately prevent children from acquiring a V2 grammar — after all, children are able to learn low frequency phenomena. In other words, even though a large proportion of the input becomes ambiguous with respect to V2, the remaining non-subject-initial sentences should provide sufficient evidence for a V2 micro-cue. The role of subject-initial sentences must therefore be more indirect. The indirect effect of subject-initial sentences derives from the vulnerability of the V2 grammar they create. That is, a V2 system becomes weakened and therefore more prone to be lost.

In summary, in all three models there is a role for the increasing frequency of subject-initial sentences in languages that have lost V2. The models interpret the rise of subject-initial sentences as a cause and not as the result of the change. Hence, the models provide a uniform explanation for the loss of V2 in that the change is connected to the nature of the evidence in the input. In this thesis, I will develop an account that shifts away from the role of unambiguous evidence in the loss of V2. Based on domain-general findings that variability in the input benefits learning, I will test how variation relates to the loss of V2. Specifically, the role of the increase of subject-initial sentences and subsequently ensuing change in variability will be examined.[86] That is, the interaction of learning and the input will still be fundamental, yet the emphasis will be placed differently. In the next section, I will outline the role of variability and spell out its relation to the loss of V2.

1.5 The role of variability in language learning and its relation to V2

All three learning models discussed in the preceding section recognise the rise of subject-initial sentences as a cause for the loss of V2. At first sight, this conclusion may seem at odds with the findings of Westergaard (2009b, 2011) that learners can still reliably acquire low frequency patterns. However, the developments in the languages reviewed in §1.3 clearly demonstrate that V2 was lost in face of the still existing evidence for V2 (albeit infrequent). In the model of Westergaard (2009b, 2021a), the diachronic developments can be explained by the effect the low frequency of unambiguous evidence has: constructions become susceptible to change. This raises the question though why constructions are prone to change when the frequency is low. For V2, I will argue that features of human cognition can be invoked as part of the explanation — in other words, a third factor explanation is offered (cf. Chomsky 2005). More precisely, I will argue that variability in the clause-initial constituent is a driving factor for the loss (or

[86] The initial hypothesis will define grammatical functions as the relevant domain of variability. Based on experimental findings in Chapter 3 and Chapter 4, however, I will conclude that variability in either grammatical functions or categories benefits the learning of a V2 language.

retention) of a V2 grammar. This will allow me to provide an explanation for the loss of V2 incorporating the rise of subject-initial sentences and learning. Such an analysis is in line with Walkden (2021a: 13) who argues that every theory of language change should incorporate properties of individual speakers (i.e. acquisition, cognition and language use). I will first develop the hypothesis connecting variability to the loss of V2 (§1.5.1) before exploring the type of variability needed for the acquisition of V2 (§1.5.2).

1.5.1 The role of variability in the loss of V2

There is strong evidence that learning in general is impacted by variability. Raviv, Lupyan & Green (2022) reviewed experimental work from a wide range of cognitive domains, ranging from motor skills to problem solving to language learning. The review revealed converging evidence across all covered domains that variability fosters learning. Variability does not constitute a uniform concept in that (at least) four different types can be distinguished: numerosity, heterogeneity, situational/contextual and training scheduling (Raviv, Lupyan & Green 2022). Numerosity denotes the number of training items that learners encounter during learning. For example, a child that is being familiarised with the concept of cheese might be shown three or twelve instantiations of cheese. Heterogeneity, in turn, describes the degree to which the examples differ. The child trying to grasp the concept of cheese with the help of visual examples might encounter cheeses whose visual appearance is rather similar such as Appenzeller, Emmental and Gouda. That is, the child would see three hard cheeses with a (light) yellow colour. A more diverse sample, however, would be Red Leicester, Blue Stilton and mozzarella exhibiting a broader range of colour as well as texture.[87] The third type of variability captures "variation in the external learning conditions" (Raviv, Lupyan & Green 2022: 468). In the context of the cheese example, situational/contextual variability would occur when the child is exposed to cheese in different locations such as at the cheese monger, the supermarket and in the kitchen at home. The final type of variability concerns the training schedule. A child might be shown exclusively Emmental when cheese is introduced for the first time and exclusively Peccorino the next time. Alternatively, a more varied familiarisation could be pursued where Peccorino and then Emmental are shown the first time. The next time the child is familiarised with cheese, the reversed order could be used, i.e. first Emmental and then Peccorino.

When it comes to learning in the context of language, a fostering effect of variability is well documented. Variability in different linguistic domains has been shown to benefit learning. Phonetic variation during learning, for instance, increases the word recognition at test both in infants (L. Singh 2008, Rost & McMurray 2010, Galle, Apfelbaum & McMurray 2015) and adults (Sommers & Barcroft 2006, Sumner 2011). Furthermore, phonemic contrasts in non-native languages become more salient after exposure to acoustic variation (Leong et al. 2018). Relatedly, the exposure to more speakers dur-

[87] Raviv, Lupyan & Green (2022: 470) observe that numerosity and heterogeneity are often confounded, because numerosity served as proxy for heterogeneity in many studies. The cheese example shows however that this does not need to be the case.

ing learning (i.e. increased talker variability) leads to better word segmentation in infants (Graf Estes & Lew-Williams 2015) and better learning of phonemic contrast in adults (Lively, Logan & Pisoni 1993). In addition, unfamiliar dialect speakers are more accurately categorised by adults after being exposed to more speakers from the same dialect (Clopper & Pisoni 2004). Native speakers also adapt better to non-native speakers when previously exposed to more non-native speakers sharing the same L1 (Bradlow & Bent 2008). Talker variability also benefits vocabulary learning in a L2 (Barcroft & Sommers 2005) even though this might only apply to adults but not children (Sinkeviciute et al. 2019).[88] Besides acoustic variation and talker variability, contextual diversity also fosters learning. When learners encounter novel words in more semantic contexts during learning, these words are learnt better (Adelman, Brown & Quesada 2006, Frances, Martin & Duñabeitia 2020, Hsiao & Nation 2018, Pagán & Nation 2019).

Even though those studies demonstrate the benefits of variation in different linguistic domains, they do not touch on the domain relevant for the goals of this thesis — syntax. In fact, work on variability in the syntactic domain is scarce. The few studies that have been conducted, however, converge on the same finding as the studies in the other linguistic and cognitive domains: variation fosters learning. Poletiek & van Schijndel (2009) investigated how much input learners need to learn structures generated by a finite-state grammar. They systematically compare the effect of numerosity and heterogeneity. Interestingly, their results suggest that learners solely benefit from more heterogeneous input. A similar conclusion is reached by Schiff et al. (2021) who addressed a related question in their study. Akin to Poletiek & van Schijndel (2009), the mere increase in the number of training items did not result in better learning. Only once the variability between the examples surpassed a certain point, learners could benefit from the variation.

Another piece of evidence for a connection between variability and learning in the syntactic domain can be found in Gómez (2002) and Gómez & Maye (2005). Both studies investigated the acquisition of non-adjacent dependencies by adults and infants. Non-adjacent dependencies are structures with two dependents that are separated by intervening material. Those dependencies may take the form of 'aXc' where 'a' elements need to be followed by elements from the 'c' category; the elements separating the two dependents are represented by 'X'. Structures of this kind are a common feature of natural languages. The examples in (77) illustrate this for English. Both, the agreement between the subject and the verb in (77a) and the *auxiliary* + *-ing* construction in (77b) can be considered non-adjacent dependencies (Gómez 2002: 431, Gómez & Maye 2005: 184).

(77) a. *The mouse* despise*s* all types of soft cheese.
b. The mouse *is* steal*ing* a large loaf of cheese.

Gómez (2002) and Gómez & Maye (2005) exposed both infants from different age groups and adults to non-adjacent dependencies of the type 'aXc' in an artificial language

[88] It is also worth noting that talker variability does not automatically entail a beneficial effect on learning as Atkinson, Kirby & Smith (2015) failed to find an effect of speaker input variability.

learning study (cf. §1.6). While the set size of 'a' and 'c' elements was kept constant, the set size of 'X' was manipulated between conditions. The amount of training items remained constant across conditions, meaning that the effect of heterogeneity was effectively assessed. Interestingly, only adults and infants (17 and 18 months old) in the condition with the highest variability in 'X' were able to learn the dependency between 'a' and 'c' (Gómez 2002, Gómez & Maye 2005).[89] Based on these findings, Gómez (2002: 435) suggests that higher variability helped participants to identify the relevant dependency relation by shifting the attention from uninformative adjacent dependencies ('aX' or 'Xc') to informative non-adjacent dependencies ('a' & 'c'). Gómez & Maye (2005: 201) propose an alternative interpretation invoking memory constraints. Learners can keep track of the frequencies of a small number of individual elements. In contexts with many intervening 'X' elements however, only 'a' and 'c' can be tracked causing learners to recognise the dependency relation.

The hitherto discussed studies demonstrate that variability benefits learning, but it has been left open why variability is beneficial in the first place. Raviv, Lupyan & Green (2022) argue that generally three mutually non-exclusive explanations can be adduced for the fostering effect of variability. First, learners can identify which features are relevant and which irrelevant for the task at hand. Hence, when learners encounter low variability (e.g. due to a non-representative sample) during training, they may not be able to generalise appropriately (Raviv, Lupyan & Green 2022: 473). This explanation is also used by Gómez (2002) to explain their experimental findings. The variability of the intervening elements helps learners to focus on the relevant aspect, the non-adjacent dependencies. Second, variability helps learners to generalise because the variability provides a broader coverage of the task or phenomenon. Learners generally fare better at interpolating than extrapolating indicated by a lower error rate and faster performance (Raviv, Lupyan & Green 2022: 475). The third and final benefit is the faster retrieval from memory (Raviv, Lupyan & Green 2022: 475).

Against the backdrop of previous work providing evidence for the benefits of variability during learning, the question arises what implications (if any) these findings have for the loss of V2. Languages that were in the process of losing their V2 grammar show a shift in the composition of the clause-initial position. As extensively discussed above, the proportion of subject-initial sentences increased, whilst the proportion of non-subjects showed the opposite trend. That is, the composition became less heterogeneous; the variability declined as subjects accounted for the vast majority of clause-initial constituents. If variability fosters learning, the lower variability should result in worse learning and eventual loss of V2. This can be formulated as the following hypothesis:

(78) **Hypothesis**
Low variability in the clause-initial constituent will lead to the loss of V2.

[89] Gómez & Maye (2005) also tested infants younger than 17 months old. While some of the 15-month-olds showed signs of learning, 12-month-olds did not — despite slight changes to the size of 'X' to accommodate their younger age. Surprisingly, adults with language-based learning disabilities also failed to acquire the non-adjacent dependency (Grunow et al. 2006) even though higher variability can indeed foster learning in this group, albeit of *adjacent* dependencies (von Koss Torkildsen et al. 2013).

This hypothesis receives independent support from speakers of heritage V2 languages. Westergaard, Lohndal & Lundquist (2023) investigated spontaneous productions of 50 speakers of Norwegian heritage language. Crucially, speakers who produced a significantly higher proportion of non-subject-initial V2 sentences displayed fewer V2 violations overall.[90] Such a correlation is expected if less variability impedes learning of the V2 grammar.

If the hypothesis is borne out, it is possible to explain *how* V2 was lost due the lack of variability. Crucially however, the hypothesis does not provide an explanation for *why* variability is beneficial in the first place. Of the three factors highlighted by Raviv, Lupyan & Green (2022), the first and the second factor (i.e. identification of relevant features and better generalisation) are arguably the most relevant ones. Learners faced with the task of acquiring a V2 grammar need to form representations where the clause-initial position does not exhibit a fixed association with a particular category. That is, learners need to abstract away from the associations between grammatical functions and the clause-initial position they encounter in the input. A high proportion of subject-initial sentences will impede the necessary generalisations. The hypothesis in (78) can therefore be revised as (79) to reflect this aspect:

(79) **Hypothesis** (*revised*)
Low variability in the clause-initial constituent will lead to the loss of V2, because learners are unable to form generalisations about the flexibility of the clause-initial position. Learners will instead stipulate a fixed association of the clause-initial position with a grammatical property.

The hypothesis remains vague with respect to the exact nature of the evidence. So far, it has only become clear that non-subjects need be involved somehow. The following subsection will elaborate on the nature of the relevant category in more detail.

1.5.2 The evidence for the acquisition of V2

In earlier work, attempts have been made to establish a universal trigger for V2. This has been influenced by the notion of parameters which have to be set by children during acquisition. According to Fodor (1998), learners need to be exposed to unambiguous structures in their input in order to set a parameter correctly. Sitaridou (2011: 164, 2012: 585), for instance, proposes that XP-V-S_{pron} structures (i.e. subject-verb inversion with pronominal subjects) are the sole trigger for V2, as subject-verb inversion with nominal subjects occurs in non-V2 grammars as well, such as Greek (80a) (cf. also Kaiser 2002: 112). Crucially however, the corresponding structures with pronominal subjects are ungrammatical, as the example in (80b) evidences.[91] An alternative proposal for an universal V2 trigger is made by Fodor (1998: 25). Independent from underlying word

[90] Additional (albeit indirect) evidence is provided by Larsson & Kinn (2022) for another heritage language. In American Swedish heritage language, the proportion of non-subject-initial V2 sentences decreased, while non-V2 and SV structures increased in frequency compared to homeland Swedish. This can be explained if variability in clause-initial position is beneficial for learning.

[91] But not in some Romance non-V2 languages, as the example in (81) highlights (cf. Varga 2017).

order properties (i.e. relative order of subject and verb as well as object and verb), a non-subject in clause-initial position followed by an auxiliary will trigger V2. However, as pointed out by Kaiser (2002: 113), such word orders also exist in non-V2 languages such as Portuguese, illustrated in (81).

(80) a. Xthes **ipevale** *i Maria* tin paretisi tis. GREEK
yesterday present-3SG.PST Mary the resignation her
'Maria handed in her resignation yesterday.'
b. ?*Xthes **ipevale** *afti* tin paretisi tis.
yesterday submit-3SG.PST she the resignation her
'Maria handed in her resignation yesterday.'
(Sitaridou 2012: 583, 585)

(81) Talves **tenha** *ela* lido o livro. EUROPEAN PORTUGUESE
maybe has she read the book
'Maybe she read the book.'
(Kaiser 2002: 113)

The inability to identify a universal structure serving as evidence for a V2 grammar casts doubt on the existence of such a structure. This matches the position Yang (2002: 35) and Clark & Roberts (1993: 319) advocate for: Grammars in general cannot be unambiguously identified by learners based on a specific structure. Instead, the evidence should be on a smaller scale (i.e. substructures) in addition to being language- and context-specific, as suggested by Westergaard (2009b). This is poignantly illustrated by the diachrony of English: While English has lost the V2 property in declaratives, *wh*-interrogatives still require a V2 word order (cf. §1.3.1). This effectively means that various aspects need to conspire to 'birth' a V2 grammar in speakers during acquisition. Confining the evidence to the clause-initial position fulfils this need. It is only with the interaction of verbs placed in the second position that the variability in the clause-initial position can affect the learning of a V2 grammar. When considered from this perspective, the postverbal word order and the nature of specific elements following the verb (e.g. pronominal subjects vs. nominal subjects) are not relevant for the V2 phenomena itself, though of course, they may be relevant for the acquisition of other aspects of the grammar.

What then is the relevant variability in the clause-initial position? It is, in principle, conceivable that learners display sensitivity to two different kinds of evidence. Learners may either be sensitive to grammatical functions (e.g. subject, direct object) or to grammatical categories (e.g. NP/DP, PP, AdvP, CP).[92] While Yang (2000, 2002) argues for the former option, Lightfoot (1999, 2006) assumes the latter. There is no inherent reason to prefer one over the other. Grammatical rules can be sensitive to grammatical functions — such as the commonly assumed movement of English subjects to the inflectional domain (i.e. IP/TP) — or grammatical (sub)categories. Object shift in the Mainland Scandinavian language can be considered evidence for the latter, as the movement is restricted to pronominal objects, cf. (82).

[92]However it should not be ruled out prematurely that learners are sensitive to both simultaneously.

(82) a. **Peter* **læste** bøgerne uden tvil aldrig. DANISH
Peter read books-the without doubt never
'Peter has, without doubt, never read books.'
b. *Peter* **læste** dem uden tvil aldrig.
Peter read them without doub never
'Peter has, without doubt, never read them.'
(Vikner 2006: 393)

I will stipulate — as a starting hypothesis — that grammatical *functions* constitute the relevant evidence. This is solely motivated by the fact that grammars are usually formulated in terms of grammatical functions (e.g SVO, SOV). In the context of the hypothesis in (79) then, this assumption entails that a low variability of grammatical functions in initial position should lead to the loss of V2. It should be pointed out that variability in this context is finite. The highest variability is reached when different grammatical functions are attested with the same frequency in clause-initial position — in other words a uniform distribution. The lowest variability in turn occurs when a single type of grammatical function accounts for all instances. It is important to re-emphasise that the focus on grammatical functions as relevant domain of variability is only supported by circumstantial evidence. As shown above, grammatical rules can be sensitive to grammatical categories as well. That is, none of the two alternatives should be inherently favoured from a theoretical perspective. In fact, as I will discuss in later chapters, the findings of this thesis do provide evidence that learners can be sensitive to variability in grammatical functions and grammatical categories: Chapter 3 suggests that grammatical categories constitute the relevant domain of variability, Chapter 4 points to the conclusion that learners can be sensitive to either grammatical categories or grammatical functions.

Even though the variability in the clause-initial constituent has been identified as the critical evidence for a V2 grammar, it does not mean this is the only type of evidence learners have at their disposal. Adams (1987b: 87) argues that the alternation between V2 and V-final order in German and Dutch functions as additional evidence for learners. Similar claims have also been made elsewhere in the literature (e.g. Roberts 1993: 157, Vance 1997, Willis 1998: 183). This explanation seems intuitive from a learning perspective. Different verb positions at opposite clause edges must be more salient to learners than alternations between SVO and V2 sentences where no or only little differences can be observed.[93] The alternation thus highlights the derived character of V2. Whether this renders learners less susceptible to the variability in the clause-initial position must be addressed in future work.

A further type of evidence that resembles the word order alternation can be found in V2 sentences. Particle verbs such as *abschneiden* 'cut off' are split in V2 clauses: The verb moves to the second position while the particle has to remain in the position

[93]Sentence adverbs and negators can in principle help to disambiguate SVO and V2 clauses (Waldmann 2008, Westergaard 2009b). This, however, is context-dependent and of course not all clauses comprise either a negator or an adverb. In Norwegian for instance, subject-initial declaratives that contain either an adverb or a negator account for 6.2% of all V2 evidence (Westergaard 2009b: 67).

where verbs are basegenerated (83a). That this is in fact the base position can be seen in the embedded clause in (83c). If the particle is pied-piped to the second position, the sentence becomes ungrammatical (83b). Akin to the alternation between V2 and V-final, the stranded particle could highlight the derived nature of the V2 word order.

(83) a. *Der Käser* **schneidet**$_1$ etwas Käse für die Maus **ab** t_1. GERMAN
the cheesemaker cuts some cheese for the mouse off
'The cheesemaker cuts off some cheese for the mouse.'
b. **Der Käser* **abschneidet**$_1$ etwas Käse für die Maus t_1.
c. …dass *der Käser* etwas Käse für die Maus **abschneidet**.
…that the cheesemaker some cheese for the mouse cut off
'that the cheesemaker cuts off some cheese for the mouse.'

Although likely contributing to the acquisition of a V2 grammar, the impact of any additional evidence will not be further examined in this thesis.[94] First and foremost, the focus here lies on the role of variability in the clause-initial constituent in the loss of V2. Hence, the inclusion of additional types of evidence for V2 could confound the results. Moreover, the alternation between V2 and V-final clauses does not constitute a feature of all V2 languages. Among the Germanic languages, the verb alternation can only be found in Dutch, Frisian and German. That is, this type of evidence would not contribute to the development of an explanation for the loss of V2 across languages.[95] Nevertheless, future work should investigate how these additional types of evidence contribute to the acquisition of V2.

Investigating the effect of variability of different grammatical functions in the clause-initial position is non-trivial. To test the hypothesis derived in this chapter experimentally, the input of child and adult learners would need to be manipulated. However, such an approach would most likely necessitate an unethical study design: Full control over the input frequencies can only be obtained when participants are kept isolated from other speakers for a prolonged period, especially in the context of L1 acquisition. Such a design must therefore be discarded and an alternative has to be sought. Recall that both Gómez (2002) and Gómez & Maye (2005) — amongst others — have relied on an artificial language learning paradigm for their experiments. That is, they have created a language suitable to their needs that was then taught to participants. Given that the shared relevance of variability in both Gómez (2002) and this thesis, an artificial language learning study appears to be a fruitful alternative to unethical designs. Besides, artificial language learning has been increasingly used to explore how features of individual learners and the input interact. What is more, this experimental paradigm has been successfully used both with adults and with children. That is, experimenters can use the same methodology for testing two different participant groups. In the present thesis then, both adults and children could be tested. However, given that adults are easier to recruit and test than children — for instance due to the availability of online

[94] This also applies to evidence *against* V2. V>2 orders likely have a negative effect on the acquisition of a V2 grammar.
[95] There are also practical arguments against the inclusion, as the discussion in §2.2 will show.

crowdsourcing platforms like Amazon Mechanical Turk or Prolific — testing adults would offer certain practical advantages. This should not affect the interpretability of the results since the results of experiments with adults and children have been shown to converge: In the aforementioned studies of Gómez (2002) and Gómez & Maye (2005) for instance, both adults and children are susceptible to the variability in the input. A further example for converging results of adults and children has been found for the word order in noun phrases (Culbertson, Smolensky & Legendre 2012, Culbertson & Newport 2015).[96] In the next section, I will therefore illuminate the artificial language learning paradigm more closely.

1.6 Artificial language learning: A (new) methodology

Variability in the input plays a crucial role during language acquisition, as the previous section has shown. Based on the observation that the variability of clause-initial constituents decreased in languages that were in the process of losing V2, I hypothesised that this decline is ultimately responsible for the loss of V2. Evidence for the role of variability during learning (in the linguistic domain) has been provided by artificial language learning studies. This methodology therefore lends itself to testing my own hypothesis experimentally. In particular, it allows me to manipulate the distribution of clause-initial constituents in the input learners receive and subsequently observe the effects of the manipulation on learning. In what follows, I will provide a brief overview of the methodology used in this thesis.

Artificial language learning (ALL) is an experimental paradigm in which participants are asked to learn a miniature novel linguistic system of varying complexity (Folia et al. 2010, Culbertson 2012, 2017, 2023, Ettlinger et al. 2016). Participants' acquired knowledge is subsequently measured using a wide variety of tests. Most ALL studies elicit behavioural data, though a range of studies has also collected physiological responses (Friederici, Steinhauer & Pfeifer 2002, Friederici et al. 2006, Morgan-Short et al. 2012). Adults and children form the typical participant groups (Gómez & Gerken 2000, Culbertson & Schuler 2019), but species other than humans are studied as well, including monkeys, apes and songbirds (Fitch & Hauser 2004, Gentner et al. 2006, Claidière et al. 2018, Watson et al. 2020). The ALL paradigm allows researchers to create a controlled environment to measure the effects that arise when a language is being learnt (Culbertson 2023): While natural languages form an intricate system of interrelated properties that could act as confounds, ALL enables the matching of relevant properties such that only the variables of interest are manipulated. Furthermore, experimenters can control the amount of input participants receive during training. An additional advantage pointed out by Culbertson (2023) is the option to test linguistic patterns that are either rare or completely unattested in the languages of the world.

ALL has been in use since Esper (1925) with designs similar to modern ones starting to emerge in the 1960s (Braine 1963, 1966, Reber 1967). In linguistics however, it was

[96] Note that this does not entail that the results of adults and children will *always* converge. Adults and children can still display diverging behaviours in experiments (Culbertson & Newport 2017, Tal & Arnon 2022).

only in the last two decades that the interest in ALL as a methodology has significantly increased. A broad range of questions have been addressed with the help of ALL, such as the role of cognitive constraints on learning and their impact on typological universals (Culbertson 2012, 2017, 2023), L2 acquisition (Rebuschat & Williams 2012, Ettlinger et al. 2016) and more recently L3 acquisition (Jensen & Westergaard 2023, Mitrofanova, Leivada & Westergaard 2023). Irrespective of the specific research question, existing work encompasses all domains of linguistics, ranging from phonology (Moreton 2008, White 2014, Yin & White 2018, Prickett 2019), morphology (Liter, Heffner & Schmitt 2017, Saldana et al. 2021, Saldana, Herce & Bickel 2022), syntax (Culbertson & Adger 2014, Culbertson & Newport 2017, Tabullo et al. 2012, Franck, Rotondi & Frauenfelder 2016) to semantics (Silvey, Kirby & Smith 2015, Saratsli, Bartell & Papafragou 2020, Maldonado & Culbertson 2022) and pragmatics (Saratsli & Papafragou 2023). Given the widespread use of ALL, it is indispensable to show that ALL and natural language learning involve the same cognitive processes. After all, artificial languages are acutely simplified compared to natural languages (Culbertson 2023) and might thus be represented differently in the minds of participants. Two arguments can be adduced in favour of the parity of artificial and natural languages. First, the results of ALL studies align with typological patterns attested in natural languages (Culbertson 2012, 2017, 2023, Culbertson & Schuler 2019). That is, the preferences shown by participants are identical to those observed in natural languages. The second argument concerns the way languages are learnt. Ettlinger et al. (2016) show that artificial languages are learnt similarly to natural languages (cf. Folia et al. 2010). The performance of learners in a L2 correlated with the their performance when learning in an artificial language. Relatedly, constructed languages elicit responses in areas of the brain that are also activated for natural languages (Malik-Moraleda et al. 2023). Hence, any concerns with regards to the validity of ALL can be dismissed.

When it comes to the design of artificial languages, two factors need to be taken into consideration independent of individual research questions. A crucial aspect of every language is the semantic space. On the one hand, referents and actions might already exist (e.g. house, kiss) such that participants do not require any familiarisation with the concepts. On the other hand, referents and actions might be unfamiliar for participants (Kirby, Cornish & Smith 2008, Culbertson, Smolensky & Legendre 2012, Culbertson et al. 2020), which is useful for avoiding any influence from participants' L1 (see below). This is typically accomplished with the use of novel objects (e.g. from Horst & Hout 2016) and atypical movement paths (e.g. up-right-down-right-down). A third alternative is an empty semantic space. This kind of semantic space is predominantly used in connection with statistical learning tasks where participants typically listen to recordings of concatenated syllables (Saffran, Aslin & Newport 1996, Aslin, Saffran & Newport 1998, Wonnacott, Brown & Nation 2017).

The second factor is related to the semantic space, namely the lexicon. In principle, three options are available: lexical items taken from natural languages, semi-artificial lexical items and fully-artificial lexical items. The first option, i.e. natural language lexical items, relies on lexical items from participants' native language (e.g. *duck*, *kick*). The elements rendering the language artificial are provided by either an alien syntax

or morphology (Smith & Wonnacott 2010, Rebuschat & Williams 2012, Culbertson & Adger 2014, Martin et al. 2019). Semi-artificial lexical items stand in an onomatopoetic (or iconic) relationship with a natural language, such as *kwako* 'duck' in Atkinson, Smith & Kirby (2018). Alternatively, the words of the artificial language show close resemblance with a word from a natural language (e.g. *kit* 'cat').[97] Fully-artificial lexical items form the final category of lexical items. This type is characterised by the complete absence of any similarity to natural languages participants are familiar with (Tabullo et al. 2012, Getz 2018). The word *duck* might thus represented as *stapi*. The choice of lexical item type is guided by the research question. If the focus lies on higher level orders or patterns, natural or semi-artificial lexical items offer two advantages. First, they allow participants to focus on the more complex orders in the input without being preoccupied with lexical learning. Second, participants can be exposed to more variability in the training patterns. This decreases the chance of confounding effects due to specific lexical items. On the other hand, variability might be one of the factors that need to be controlled in the experiment (cf. §1.5). Crucially, experiments using natural language lexical items and fully-artificial lexical items show converging results (Culbertson & Adger 2014, Martin et al. 2020).

Despite the significant advantages that accompany ALL, experimenters need to approach this methodology with caution nonetheless. First and foremost, learners (both adults and children) cannot be construed as tabula rasa — prior linguistic knowledge does affect learners' performance during a learning task (Siegelman et al. 2018, cf. also Hamrick & Sachs 2018). Importantly, transfer effects are not confined to a particular linguistic domain but have been noted for phonotactics (Finn & Hudson Kam 2008), syllables (Elazar et al. 2022), the lexicon (Tang & Baer-Henney 2023) as well as underlying word order (Onnis & Thiessen 2013).[98] However, the dissociation of previously acquired languages and the artificial language can be achieved through careful design choices. On the one hand, one might test a group of participants whose previously acquired languages do not exhibit the feature of interest. When investigating V2, this option offers itself as the preferred option. In particular English speakers constitute a group of speakers that is relatively easy to access without exhibiting a V2 grammar. Alternatively, different populations could be tested and subsequently compared. Two general options are available here: Either, multiple populations who share certain features are tested

[97] Note that artificial languages combing lexical items from natural languages with alien elements such as a word order deviating from the source language are also referred to as semi-artificial languages. This applies to the languages used in Chapter 2 and Chapter 3.

[98] Although this may pose challenges for experimenters in terms of the experimental design, the effect of prior linguistic knowledge can be considered as further evidence for the suitability of ALL to study questions in relation to language. If ALL and natural language learning did not fundamentally rely on the same cognitive processes, transfer would likely not occur. This is further underscored by the existence of transfer effects in natural L2 and L3/L*n* acquisition (Westergaard 2021b). Such parallels between the acquisition of natural and artificial languages are not expected to occur if the underlying processes are not identical. This can be taken as evidence for the ecological validity of ALL.

(Culbertson et al. 2020), or populations with opposing features are compared (Martin et al. 2019, cf. Culbertson & Adger 2014).[99]

A potentially problematic aspect with regards to the goals of this thesis can be seen in the range of phenomena that have been investigated using ALL. Focusing on morphological and syntactic phenomena, most studies are concerned with 'smaller' linguistic structures; in other words, phenomena that do not include the whole domain but only subparts of it. With regards to morphology, this includes the preference for suffixes versus prefixes (Hupp, Sloutsky & Culicover 2009, Martin & Culbertson 2020), person-number agreement (Saldana, Herce & Bickel 2022) and the order of different morphemes (Saldana, Oseki & Culbertson 2021). As for syntactic phenomena, much work has centred around the ordering of modifiers in the noun phrase (e.g Culbertson, Smolensky & Legendre 2012, Culbertson & Newport 2015, 2017, Martin et al. 2020, Saldana et al. 2021) and basic word order (Goldin-Meadow et al. 2008, Schouwstra & de Swart 2014, Tily, Frank & Jaeger 2011, Tabullo et al. 2012, Fedzechkina, Jaeger & Newport 2012, Fedzechkina & Jaeger 2020, Kirton et al. 2021). Given that V2 is a complex word order derived through application of various processes (cf. §1.2.3), the question arises whether ALL is in fact suitable for studying questions related to V2. I am, however, confident that the answer is yes, for two reasons. First, several studies have demonstrated that simple sentences involving subjects, objects and verbs can be learnt in an ALL experiment (Tabullo et al. 2012, Fedzechkina, Jaeger & Newport 2012, Fedzechkina, Newport & Jaeger 2017, Roberts & Fedzechkina 2018, Fedzechkina & Jaeger 2020, Tal & Arnon 2022, Tal et al. 2022).[100] That is, at least simple clause structures can be learnt in a relatively short time period. Second, a small number of ALL studies have used a V2 language even though V2 did not form the object of study itself (Rebuschat 2008, Rebuschat & Williams 2012, Getz 2018, Gao & Ma 2021). Crucially, participants in those experiments were able to learn the input structures. Notwithstanding this finding, issues remain with the experimental design and the way learning was measured in these V2 experiments. I will therefore develop an experimental paradigm in Chapter 2, informed by previous studies, that will allow me to test the hypothesis developed in §1.5.

1.7 Summary

This chapter introduced the verb second (V2) phenomenon which is characterised by the obligatory placement of the finite verb in the second position of the clause and the virtual absence of any constraints on the single constituent preceding the verb.

[99] Depending on the research question, a change of the modality might also constitute a viable option. Silent gesture has been increasingly used as alternative for language in the spoken and written modality (Goldin-Meadow et al. 2008, Schouwstra & de Swart 2014, Motamedi et al. 2019, Kirton et al. 2021). In this paradigm, non-signing participants are asked to express events solely using their hands forming gestures. As pointed out by Culbertson (2023) however, the relation between linguistic structure and silent gesture is opaque — at least for the time being.

[100] This list does not include studies where participants were taught more complex strings void of meaning or at least semantically very impoverished (Morgan, Meier & Newport 1987, 1989, Reeder, Newport & Aslin 2013, 2017).

V2 is commonly analysed as derived phenomenon. That is, the position of the clause-initial constituent and the finite verb are taken to be the result of movement to the left periphery of the clause. V2 is typologically exceptionally rare and mostly confined to the Germanic languages (with the exception of English). Most of the Romance languages in addition to English and Welsh used to be V2 languages but have subsequently lost their V2 property. Earlier stages of the languages which lost their V2 grammar exhibited many parallels with earlier stages of Germanic V2 languages: generalised V-to-C movement and variability in XP-movement to the left periphery. Whilst the Germanic languages have established a more rigid V2 system, the Romance languages, English and Welsh underwent change into the opposite direction. This then poses the question why diverging developments occurred. After reviewing at least some explanations for the loss of V2 in individual languages, the drivers for the loss of V2 appear to be language-specific. However, it is also evident that the various accounts share important insights: First, learning is purportedly playing a role; second, the diachronic developments show an increase in subject-initial sentences. This forms the starting point of the present thesis: The goal is to develop an account that provides an explanation for the loss of V2 across languages, building on a connection between learning and an increased frequency of subject-initial sentences. For this purpose, three models connecting language change with learning were examined. All three theories emphasise the importance of the evidence in the input. This was then examined from a cognitive context in terms of variability. Various studies have found converging evidence for the beneficial effect of variability for learning, including language learning. These findings served as foundation for the hypothesis that will be tested in this thesis. The loss of V2 is hypothesised to be conditioned on the variability in the clause-initial position. Less variability should inhibit participants' ability to generalise the lack of constraints on elements in the clause-initial position which then leads to the loss of V2. Grammatical functions were stipulated to be the relevant domain of variability (although this assumption requires revisions in later chapters). I also introduced the methodology that will be used in the present thesis, namely artificial language learning.

This thesis is structured as follows: Chapter 2 reviews existing ALL studies that have used artificial V2 languages. Based on a critical discussion of their design, a new design is developed and put to test in two experiments using a semi-artificial language. A key insight of the experiments is the combination of reading and production trials during training. Moreover, Prolific users are identified as target sample. This design is then used in Chapter 3 to investigate the hypothesis derived in the present chapter. Three different languages are compared, namely a maximally variable language (i.e. uniformly distributed clause-initial constituents) and two languages with a similar degree of variability but different dominating constituent types (i.e. objects and adjuncts). The results show that a language where adjuncts dominate the clause-initial position is learnt best, followed by the uniform language. Moreover, the results of a large-scale corpus study on the distribution of clause-initial constituents in German are reported. Although subjects account for the majority of all clause-initial elements, adjuncts occur with considerable frequency in clause-initial position. The corpus results suggest that a significant proportion of clause-initial adjuncts facilitate the acquisition of a V2 grammar.

In order to differentiate between two possible explanations for the experimental results of Chapter 3 — confounding influence of participants' L1 and a genuine learning advantage — the results of two further experiments are discussed in Chapter 4. Instead of using a semi-artificial language, a fully-artificial and a visual artificial language that expresses words with icons, are used. While no learning was found for the fully-artificial language, the results of Chapter 3 could partially be replicated with the visual artificial language. The uniform language and the adjunct-dominant language were learnt better than the language where objects dominate the initial position. Crucially, no difference between the uniform language and the adjunct-dominant language could be observed. This finding is interpreted as evidence for the interference from participants' L1 in the experiment with the semi-artificial language. At the same time, the results also suggest that learners can be sensitive to variability in grammatical functions and grammatical categories. A high proportion of clause-initial adjuncts will always increase both types of variability thereby conferring a learning advantage on learners. Finally in Chapter 5, the findings of this thesis are discussed with respect to different aspects, including methodology, future work and the connection to the loss of V2 and language change more generally.

CHAPTER 2

STUDYING V2 PHENOMENA WITH ARTIFICIAL LANGUAGE LEARNING EXPERIMENTS

2.1 Introduction

In the previous chapter, I reviewed the existing literature on the diachrony of V2. Two main developments were identified: While languages like German and Dutch have been characterised by a continuous V2 grammar throughout their history, other languages have lost their V2 grammar shown at an earlier stage. Examples for the latter type include most Romance languages but also English. The question that then emerges is why some languages have retained their V2 grammar, whereas others lost it. According to the reviewed literature, the reasons for the change are manifold and language-specific. The developments, however, do share similarities. Subject-initial clauses become the clear majority pattern — at the cost of non-subject-initial clauses which decrease significantly in frequency. These shifting distributions have been used by three learning models (Yang 2002, Lightfoot 1999, Westergaard 2009b) to provide a more uniform explanation for the loss of V2. These models underscore the role of learning, specifically in interaction with the input. Even though the specifics of the models vary, they share the insight that fewer non-subject-initial clauses in the input render a V2 grammar susceptible to change.

At first sight, such a conclusion seems to be in conflict with the findings of Westergaard (2009b) that children generally reproduce low frequency patterns found in the input in a target-like way from early on. If children can acquire structures that occur with minimal frequency in the input, they should not struggle with identifying that the language in their input is a V2 language — even if non-subject-initial sentences (i.e. the evidence for V2) are scarcely attested. The diachronic developments in e.g. English and French evidently show however that this is not the case. This prompts the question why low frequency phenomena become susceptible to change as argued by the models? Work in human cognition has a potential answer. Both in linguistic and non-linguistic domains, studies have unearthed evidence that variability in the input fosters learning (Raviv, Lupyan & Green 2022). This advantage is attributed to a better identification of the features that are relevant for the learning task (as opposed to those that are irrelevant),

better generalisation due to a broader coverage of the learning task and faster retrieval from memory. Applying this idea to V2, I hypothesised that a higher proportion of non-subject-initial sentences should be beneficial for learning and consequently for the retention of a V2 grammar in a population. Encountering a high number of different types of clause-initial constituents should lead participants to the analysis that no fixed association of the clause-initial position with a particular constituent type (e.g. subjects or NPs, as I will later discuss) exists. Low variability of clause-initial constituent types on the other hand should cause participants to favour such a fixed association. The main goal of this thesis is therefore to investigate this hypothesis.

Testing this hypothesis experimentally is non-trivial as the input of learners of natural V2 languages cannot (realistically) be manipulated.[1] Fortunately, artificial language learning (ALL) offers a suitable alternative for addressing this question experimentally. Extensive previous work has demonstrated the suitability of ALL for the study of questions related to learning (e.g. Culbertson & Schuler 2019, Culbertson 2023). However, the primary focus of previous work has lain on phenomena that are less intricate than V2, such as the order inside DPs or basic word order. As a result, it is indispensable to establish first that V2 can actually be learnt in a short ALL experiment before the actual hypothesis can be put to test. A fundamental prerequisite must be that adults can learn an artificial V2 grammar sufficiently. A small number of studies have already used V2 in an ALL experiment. In this chapter, I will build on those studies to develop an experimental design that is suitable for the question at hand.

The chapter is structured as follows. I will first summarise and discuss previous artificial language learning experiments that have incorporated V2 in their design (§2.2). I will then present two experiments (§2.3 & §2.4) before discussing their results in §2.5. The main findings will be summarised in §2.6.

2.2 Previous artificial language learning studies with V2

In previous work, artificial language learning (ALL) has been established as a useful experimental paradigm for questions pertaining to learning. This suggests that ALL should also constitute a suitable methodology for the question investigated in the present thesis. To date in syntax, ALL has been primarily employed to investigate questions concerning morphosyntactic phenomena, basic word orders and word order patterns in smaller domains such as DPs (Culbertson 2023). Questions directly relating to V2 have not been addressed. There are, however, a series of studies which have used artificial V2 languages to address questions unrelated to V2. Two different strands can be distinguished. The first type are studies that used semi-artificial V2 languages.

[1] Needless to say, one could pursue non-experimental avenues. One possibility would be to investigate in corpora whether the frequencies of different types of clause-initial constituents in child-directed speech correlate with different developments in the acquisition of V2 by children. Alternatively, an approach similar to Lundquist et al. (2020) and Westergaard, Lohndal & Lundquist (2023) could be adopted who examine heritage V2 speakers heavily influenced by a non-V2 language such as English. These options will not be further pursued in this thesis as the goal is an experimental investigation. They must, therefore, be left open for further research.

That is, in these experiments, lexical items from a natural language were combined with a non-native V2 grammar. Studies of this type all build on Rebuschat (2008) and are primarily concerned with the study of L2 acquisition. The second type of ALL studies used a fully-artificial V2 language instead. Fewer studies have used this type of language, mainly Getz (2018) who examined the acquisition of morphosyntactic dependencies. In what follows, I will first provide an overview of both types (§2.2.1 and §2.2.2, respectively) before discussing how those previous designs can inform experimental designs for the study of V2 (§2.2.3).

2.2.1 Studies with semi-artificial V2 languages

To study the acquisition of syntactic rules by L2-learners, Rebuschat (2008)[2] developed an experimental paradigm that was subsequently also being adopted by other scholars (e.g. Tagarelli, Mota & Rebuschat 2015, Kim & Godfroid 2019). Specifically, the purpose of this paradigm was to scrutinise the type of syntactic knowledge (i.e. implicit or explicit) learners acquire whilst remaining unaware of the training situation (i.e. incidental learning). In this paradigm, participants are exposed to a language that consists of English vocabulary and a German word order.[3] Participants are trained on three different word order patterns: The first type (V2) is a monoclausal construction with either a subject or temporal adjunct in clause-initial position (1a). The second type (V2-VF) is a biclausal structure consisting of a main clause followed by a verb-final subordinate clause (1b). The third type (VF-V1) is also biclausal.[4] However, unlike the second type, the subordinate clause precedes the main clause (1c).[5]

(1) a. In the evening **ate** *Rose* excellent dessert at a restaurant.
b. *George* **repeated** today [that the movers his furniture scratched].
c. [Since his teacher criticism voiced], **put** *Chris* more effort into his homework.
(Rebuschat 2008: 82)

To prevent participants from becoming aware of the training situation, participants were informed they had to judge the semantic plausibility of sentences. For this purpose, half of the training materials were semantically plausible while the other half was semantically implausible. Each sentence was presented aurally. Participants had to repeat the sentence before providing their plausibility judgements. If participants judged the plausibility of sentences incorrectly, the trial was repeated.

After completing the training, participants were informed that the word order of the previously heard sentences followed specific rules. Participants were then asked to judge the acceptability of various sentences. For each of the three sentence types,

[2] See also Rebuschat & Williams (2012).
[3] Importantly, none of the participants had any knowledge of German or any other V2 language.
[4] In sentences where the clause-initial position is occupied by a non-clausal constituent, subjects and adjuncts account for 50% each. In training sentences with a clausal clause-initial constituent, adverbial clauses are the only constituent type (Rebuschat 2008: Appendix).
[5] The first experiment of Rebuschat (2008) and Rebuschat & Williams (2012) used one further sentence type. In addition to simple V2 clauses with synthetic verb forms, an additional V2 clause with an analytic verb form was included. This training condition was dropped in later experiments.

corresponding ungrammatical sentences were included with varying verb positions: Grammatical V2 sentences (1a) were contrasted with ungrammatical verb first (V1), verb third (V3), verb fourth (V4) and verb final (VF) sentences. The ungrammatical equivalent of V2-VF sentences (1b) were V1 main clauses followed by embedded VF clauses (i.e. V1-VF). VF-V1 sentences (1c) were compared to ungrammatical sentences in which embedded VF clauses preceded V2 main clauses (i.e. VF-V2). Crucially, no lexical items were reused except for some function words (Rebuschat 2008: 86). In addition to acceptability ratings, participants also had to report the confidence in their judgements as well as the basis for their judgements.

As mentioned at the outset, the design developed by Rebuschat (2008) has been adopted and modified in further work. Tagarelli, Mota & Rebuschat (2015), Tagarelli et al. (2016) and Ruiz, Tagarelli & Rebuschat (2018) included an INTENTIONAL condition (in addition to the aforementioned INCIDENTAL condition). In this condition, participants were informed at the start that the language they would encounter follows specific rules. Participants were instructed to discover these rules as they would be tested afterwards. Ruiz, Tagarelli & Rebuschat (2018) opted for a slightly different design by presenting the rules to participants and requiring them to write down two examples for each rule. The materials and other procedures remained otherwise unaltered with the exception of Ruiz, Tagarelli & Rebuschat (2018) who included pseudowords for referents into the training materials.[6] The focus of previous work on English speakers prompted Gao & Ma (2021) to replicate the study of Tagarelli et al. (2016) with Mandarin speakers. Expect for a small number of changes to the experimental materials to accommodate for participants' diverging cultural background, translations of the original materials were used. Gao & Ma (2021) also included a production test. Participants were provided with ten different expressions which they could use to construct sentences from. The only study that included a production task for English speakers is Ruiz, Tagarelli & Rebuschat (2018). Unlike Gao & Ma (2021) however, no words were provided and instead, participants saw pictures of the referents seen during training.

More significant changes to the experimental design were made by Kim & Godfroid (2019) and Miller & Godfroid (2020). While participants were emotionally induced in four different conditions (NEUTRAL, POSITIVE, NEGATIVE, COMPARISON) before commencing the language learning task in Miller & Godfroid (2020), Kim & Godfroid (2019) manipulated the manner of presentation of training and test items. In their first experiment, Kim & Godfroid (2019) compared the learning outcomes of participants who were exposed to the language in a rapid serial visual representation (RSVR) task (i.e. sentences were presented in written form, word-by-word) to those who were exposed aurally as in Rebuschat (2008). In a follow-up experiment, Kim & Godfroid (2019) again compared learners in the aural condition to participants who were exposed to the language in written form. This time, however, whole sentences were visible during training. At test, testing items were presented using both presentation manners from

[6] In a further deviation from the original study of Rebuschat (2008), Tagarelli, Mota & Rebuschat (2015), Tagarelli et al. (2016) and Ruiz, Tagarelli & Rebuschat (2018) also included memory tests. Note that different memory testes were used across the three studies. The results of those tests will not be further discussed since they bear no relevance to the question at hand.

training. That is, in experiment 1, half of the testing items were presented aurally while the other half was presented in RSVR mode. Similarly, in experiment 2, half of the testing items were presented aurally while the other half was presented in written form.[7]

The fact that multiple studies have utilised a similar experiment design enables an interesting comparison between the different studies. Unfortunately, the level of detail in the reported judgement data varies across these studies: While all studies report accuracy rates for the whole experiment, only few discuss actual acceptability rates. Accuracy rates reflect if participants have judged sentences correctly, that is if grammatical sentences were categorised as grammatical and ungrammatical sentences as ungrammatical. Acceptability rates on the other hand reflect whether participants perceive a sentence as acceptable or grammatical. Table 2.1 summarises the reported accuracy scores as well as acceptability rates for grammatical and ungrammatical structures, if provided. All studies are united by generally low accuracy scores. Two exceptions arise in that participants in the INTENTIONAL conditions in Tagarelli, Mota & Rebuschat (2015) and Tagarelli et al. (2016) exhibit somewhat higher accuracy scores. The accuracy rates of all conditions are above chance with the notable exception of the incidental group in Gao & Ma (2021). This is particular intriguing given the different language background of participants. Furthermore, participants in incidental conditions were overall less accurate in their judgements than participants in intentional conditions.

Although accuracy scores are indicative of participants' judgements, aggregating data may obscure crucial information. This can clearly be seen when ratings for grammatical and ungrammatical sentences are compared. Acceptability rates for the former are significantly higher than those for the latter (Rebuschat 2008: 93, Tagarelli, Mota & Rebuschat 2015: 234, Ruiz, Tagarelli & Rebuschat 2018: 5). This indicates that ratings of grammatical sentences were driving the observed accuracy scores. It should be pointed out, however, that the acceptability ratings are still aggregated. The only in-depth analysis of different sentence types can be found in Rebuschat (2008), reproduced in Table 2.2. The closer inspection of individual sentence types reveals that participants excelled at identifying relevant verb positions. Acceptability rates for V2 sentences were significantly higher than those for V1, V3 and V4 sentences (Rebuschat 2008: 94). Interestingly, however, Rebuschat (2008: 94) found no statistical difference between grammatical V2 sentences and ungrammatical V-final sentences. That is, participants could not discriminate between the correct verb position in main and embedded clauses. Problems with identifying the correct verb position became also visible in the ratings for grammatical V2-VF and ungrammatical V1-VF which did not differ significantly (Rebuschat 2008: 94). Again, participants could not distinguish between the correct and incorrect verb placement. On the other hand, participants were surprisingly significantly

[7] Another study that has adopted the experiment design of Rebuschat (2008) can be found in Bell (2017). Here, participants were asked to read two short stories and to complete two crossword puzzles. However, Bell (2017) did not actually use a V2 language — all V2 sentences featured adjuncts in clause-initial position. It is doubtful if this can be considered a genuine V2 language. The results will thus not be further discussed.

Author	Condition	Accuracy	Accept.gram.	Accept.ungram.
Reb(08)	INCIDENTAL	61.6	71	47
Tag(15)	INCIDENTAL	58.9	71.7	54.0
	INTENTIONAL	71.2	75.9	33.4
Tag(16)	INCIDENTAL	55.53	n/a	n/a
	INTENTIONAL	67.33	n/a	n/a
Rui(18)	INCIDENTAL	56	60	chance
	INTENTIONAL	57	65	chance
K&G(19)	AURAL1	55.63	n/a	n/a
	WRITTEN1	57.48	n/a	n/a
	AURAL2	56.53	n/a	n/a
	WRITTEN2	58.20	n/a	n/a
M&G(20)	COMPARISON	52.54	n/a	n/a
	POSITIVE	54.94	n/a	n/a
	NEGATIVE	54.45	n/a	n/a
	NEUTRAL	53.39	n/a	n/a
G&M(21)	INCIDENTAL	52.44	n/a	n/a
	INSTRUCTED	63.84	n/a	n/a

Table 2.1: Results of studies using a semi-artificial V2 language in percent, shown by conditions. Authors are abbreviated as follows: Reb(08) = Rebuschat (2008), Tag(15) = Tagarelli, Mota & Rebuschat (2015), Tag(16) = Tagarelli et al. (2016), Tagarelli et al. (2016), Rui(18) = Ruiz, Tagarelli & Rebuschat (2018), K&G(19) = Kim & Godfroid (2019), M&G(20) = Miller & Godfroid (2020), G&M(21) = Gao & Ma (2021). All studies report accuracy scores indicating the proportion of correctly identified sentences (i.e. either as grammatical or ungrammatical). Across the board, accuracy scores are overall relatively low. In addition to accuracy scores, three studies also reported acceptability rates for grammatical and ungrammatical sentences. The aggregated acceptability rates show that participants were good at identifying grammatical sentences but failed to reject ungrammatical sentences with the exception of the INTENTIONAL condition in Tagarelli, Mota & Rebuschat (2015).

Language	*V1	V2	*V3	*V4	*VF	V2-VF	*V1-VF	VF-V1	*VF-V2
semi-artif.	38.6	73.4	22.9	51.4	78.6	71.4	58.6	66.9	34.3
fully-artif.	10.7	63.3	45.3	57.3	49.3	60.2	20.0	87.3	53.3

Table 2.2: Acceptability rates of individual sentence types of the semi-artificial and artificial experiments in Rebuschat (2008) in percent. The star indicates ungrammatical sentence types. Monoclausal V2 sentences were compared to ungrammatical V1, V3, V4 and VF sentences, biclausal V2-VF and VF-V1 were contrasted with V1-VF and VF-V2, respectively. Only rates of the experimental conditions are reported here.

better at discriminating VF-V1 and VF-V2 sentences. Rebuschat (2008: 101) hypothesises that this pattern is the result of the stipulation of a micro-rule by participants. According to this micro-rule two verbs can appear sentence-medially. Taken together, the findings of Rebuschat (2008) indicate that participants were in principle able to learn that verbs are placed in second position but failed to learn the contingency of verb position and clause type.

Before turning to the fully-artificial language design, the results of the production tasks in Ruiz, Tagarelli & Rebuschat (2018) and Gao & Ma (2021) will be briefly discussed. Akin to the judgement data, participants in the instructed conditions performed better in both studies when it comes to the correct placement of the verb. This, however, does not translate to a good performance. In Ruiz, Tagarelli & Rebuschat (2018), 26% of productions of the intentional condition and 44% of the productions of the instructed condition follow the word order pattern of the language seen during training. A similar picture arises for participants in Gao & Ma (2021: 368): Participants in the incidental condition produced 14% correct sentences, while participants in the instructed condition produced 55% correct sentences. Across both studies, the majority of all correctly produced sentences are simple monoclausal V2 sentences. The production data demonstrate that participants' struggle with the position of the verb is not confined to judgements but can be observed with productions as well. This constitutes an important insight for the design, as will be discussed below (§2.2.3). I will now turn to studies using a fully-artificial V2 language.

2.2.2 Studies with fully-artificial V2 languages

In the majority of all ALL experiments with V2, participants were taught a semi-artificial language. There are, however, two studies that have used a fully-artificial design. One of them can be found in Rebuschat (2008) (experiment 4). The design was identical to the experiment of Rebuschat (2008) described in §2.2.1, the only difference being that English vocabulary items were replaced with nonsense CV syllables. This change was implemented to examine whether the use of English vocabulary affected the acquisition of the verb position negatively. Six distinct categories of syllables were created. Each

syllable category comprised two to three syllables sharing the same initial consonant (e.g. *fa*, *fi*). Unlike the previous experiments with English vocabulary, no meaning was given to the syllable sequences. Hence, the plausibility of sentences could not be probed. After training, participants had to decide for 60 syllable sequences whether they followed the patterns heard during the training. Although no novel syllables were introduced, syllable sequences used for testing had not been previously heard during training. The acceptability rates for grammatical (70.4%) and ungrammatical sentences (39.3%) did not differ significantly between the semi-artificial and artificial version of the experiment (Rebuschat 2008: 112). When individual sentence types were considered, it became apparent that participants in the fully-artificial experiment still struggled to learn the contingency of the verb position on the clause type especially in simple V2 sentences (cf. Table 2.2).[8] There was, however, some improvement over the semi-artificial experiment, as the lower rates for VF-sentences and some ungrammatical biclausal constructions suggest. Rebuschat (2008: 119) concludes that prior experience with the language does not affect the learning outcome negatively.

The second study that has relied on a fully-artificial language can be found in Getz (2018).[9] The goal of this study was three-fold. First, Getz (2018) investigated whether a complex pattern can be learnt when semantic and pragmatic information are parsimonious and the input very limited. Second, it was scrutinised whether this pattern is learnable without "full knowledge of the language's structure" (Getz 2018: 18). The final objective was to determine how the salience of closed-class elements affect the acquisition. The lexicon of the language was minimal: three nouns, two verbs, two adverbs and one inflectional marker (*kuh*) were taught to participants. Akin to Rebuschat (2008) and subsequent work, verbs could occupy two complementary positions in Getz (2018). However, the structure was simplified in that the verbal position was contingent on the presence or absence of the inflectional marker. If the verb bore the inflection marker, the verb came in second position. On the other hand, if the verb was bare, the verb occupied the clause-final position. In V-final clauses, the sentences followed the pattern S(Adv)OV, whereas V2 sentences exhibited either SV(Adv)O, OVS(Adv) or AdvVSO. During training, participants first listened to sentences before repeating them afterwards. Simultaneously, a video was played to participants showing an animal puppet performing an action (either hugging or head-butting) on another puppet.[10] After

[8]Note that the notion of verb (and of any other category) must be conceived as idealisation in this design. Given the absence of any meaning, participants could not have deduced that the category in the second position represents verbs. Rather, participants would have learnt the distributional properties of a particular semantically-undefined category of syllables. There are no inherent reasons why only verbs should be able to occur in the second position. In fact, the existence of other second-position phenomena supports such a view (Roberts 2012, Bošković 2020).

[9]See also Getz (2019).

[10]There were 38 individual sentences in the training set (26 V2 sentences & 12 V-final sentences). According to Getz (2018: 21), 53% of all sentences were subject-initial. Assuming that this figure includes V-final sentences which were exclusively subject-initial, eight of the V2 sentences were subject-initial. Consequently, the remaining V2 sentences should comprise twelve object-initial and six adjunct-initial sentences (Getz 2018: 21 fn.5). As for the distribution of V2 sentences only then, 30.77% of the sentences were subject-initial, 46.15% object-initial and 23.08% adjunct-initial.

completing the training, participants were tested on their knowledge of the language with a two-alternative forced choice (2AFC) task. For this purpose, two new puppets were introduced, who each described the same video but used different word orders. The described videos were identical to the ones participants saw during training.

Getz (2018) examined different aspects of the participants' knowledge. As for the placement of the verb, participants were overall accurate (63%). However, when different sentence types were considered individually, it became clear that this effect was driven by a high rejection rate of V1 sentences. Participants accepted a high amount of V3 sentences, thus paralleling earlier findings of Rebuschat (2008).[11] Note that this test did not only test for V2 sentences but also for V-final sentences. Apart from verb placement, Getz (2018) also tested participants' command of inflectional patterns. Participant performed well (75%) and no significant differences could be observed between sentence types. Besides inflection, Getz (2018) also tested the acquisition of fronted adverbs and objects. Akin to inflection, participants performed well (78%) and no clear differences between fronted adverbs and objects arose. The last aspect examined by Getz (2018) was participants' knowledge of word orders within sentences. Interestingly, participants performed at chance level when aggregated rates are examined. For individual rules, a more complicated picture arose. On the one hand, participants identified the correct position of postverbal objects and subjects reliably (approx. 70%, see fn.11). On the other hand, the accuracy for the postverbal order of objects and adverbs was low (approx. 30%). Accuracy scores for postverbal adverbs and subjects as well as preverbal objects and adverbs in V-final sentences were at chance level. To summarise, V2 was in principle learnt, yet participants somewhat struggled with the verb position and aspects of the word order following the verb.

Getz (2018) run three more iterations of the experiment where the inflectional marker *kuh* was either borne by the verb in final position, replaced by a phonologically heavy marker or completely dropped. The most relevant difference to the first experiment was that participants performed worse for the fronting rules. That is, participants were less accurate in picking up the correct placement of the sentence-initial constituent in V2 sentences. Getz (2018) thus concludes that closed-class markers foster the acquisition of fronting rules in V2 languages.[12] In the next section, I will discuss the implications of both the semi-artificial and fully-artificial language learning studies sketched here for studying V2.

2.2.3 Implications for the experimental study of V2

Questions surrounding the loss of V2 — and especially its hypothesised connection to the distribution of clause-initial elements (be it of grammatical functions, as hypothesised in Chapter 1, or of grammatical categories as later experiments will show) — have not been investigated with artificial language learning experiments yet. As the two preceding

[11] Unfortunately, no exact numbers are provided but based on Figure 2.3 provided by Getz (2018: 31), approximately 45% of all V3 sentences were accepted.

[12] It should be noted that the sample size in Getz's (2018) study was rather small with only 8 participants per experiment.

sections demonstrate however, a series of ALL experiments have been conducted in which participants were taught a V2 language. These studies can inform the development of an experimental design that allows addressing research questions directly related to V2. One of the central findings of previous work is the demonstration that ALL is indeed suitable for studying V2. In all of the aforementioned studies, V2 was generally learnt by participants, even though inconsistencies with respect to the position of the verb have been observed. Importantly, the nature of the language — that is semi-artificial or artificial — does not cause the miniature V2 grammar to become unlearnable in an experimental setting. By demonstrating that V2 is in principle learnable, an important prerequisite is met. Additionally, only brief exposure (i.e. approximately 30 to 40 minutes of training) is needed to learn a V2 language to a sufficient degree. This has important implications for future ALL experiments on V2. On the one hand, fewer temporal constraints apply. Albeit ALL experiments haven been successfully conducted over multiple days with both children and adults (e.g. Hudson Kam & Newport 2005, 2009), prolonged experiments might face a higher risk of participants losing interest or even dropping out prematurely (especially in an online setting). On the other hand, shorter experiments mean that financial limitations are less likely to impact data collection. This is particularly relevant when it comes to considerations of statistical power (i.e. the likelihood of correctly rejecting the null hypothesis in a frequentist framework). Without a sufficiently sized sample, studies are more likely to be underpowered — that is the likelihood of correctly rejecting the null hypothesis is low. This can have unwanted consequences for the interpretation of the results and theory building (Bergmann et al. 2018, Vasishth & Gelman 2021). In other words, shorter experiments enable the recruitment of more participants, alleviating concerns of the interpretability of the results, while also increasing the chances of retaining more participants.

A further valuable insight gained by previous work concerns the training and testing modalities. While the presentation modes of training items did not have a significant effect on the results in Kim & Godfroid (2019), the testing modes did, at least for a semi-artificial V2 language. The results indicate that a rapid serial visual representation (RSVR) testing mode should be avoided as a measurement of learning considering the observed lower accuracy scores. Although aural and written presentations of training items did not affect learning differently, the generation of written stimuli is arguably easier and prevents potential confounds from arising, for example through intonation patterns in speech recordings.

Despite the compelling evidence that V2 can indeed be learnt in an ALL experiment, issues remain with the previous experimental designs, in particular with the semi-artificial language learning experiments. In the remainder of this section I will discuss four aspects that should be taken into consideration when investigating V2 with ALL experiments. The first aspect concerns the manner in which participants are taught the language. One of the primary objectives has to be the maximisation of the learning outcome if V2 itself is the object of study. In the light of the semi-artificial experiments, this would suggest that experiments should comprise an intentional condition given the demonstrated learning advantage for participants in that condition. However, there are

good reasons to discard an intentional condition. On the one hand, instructed language learning does not reflect the processes that are usually proposed to drive language change, either with children (Lightfoot 1999, Hróarsdóttir 2003, van Gelderen 2011) or adults (Bybee 2006, 2009) as agents of language change. On the other hand, different cognitive processes could be at play when participants are tested. In a typical judgement task, explicitly taught rules might guide participants in the instructed conditions, whereas learners in the incidental condition can only rely on their intuition (and the cognitive processes affecting them, cf. Schütze 2016). Consequently, when using explicit instruction one might not measure how well a grammar has actually been learnt. Taking these two points together, the inclusion of conditions in which participants are explicitly instructed in the grammatical rule of the language, should be avoided. Instead participants should learn the language in a more incidental fashion as in Rebuschat (2008) and Getz (2018).

A second related aspect requiring further elaboration is the manipulation of plausibility. Half of all training sentences were semantically implausible in the work of Rebuschat (2008) and colleagues. While this was a necessary means of deception to simulate incidental learning in those studies, semantic plausibility has been shown to impact processing, both in children (Valian, Prasada & Scarpa 2006, Wallan 2018, Polišenská et al. 2021) and adults (Traxler & Pickering 1996, Williams, Möbius & Kim 2001, Roberts & Felser 2011, Lee & Witzel 2023). Crucially for the goal of the present chapter, L2 learners showed difficulties in recovering from misanalyses during parsing (Williams, Möbius & Kim 2001, Roberts & Felser 2011, Lee & Witzel 2023). If learners form inaccurate representations of their training input due the semantic implausibility, learning might be negatively affected and slowed down. Furthermore, for the purposes of the present thesis, emulating incidental learning conditions to the same extent as in Rebuschat (2008) can be neglected. Previous work on ALL has shown that ALL is a suitable methodology to study the effects of learning without additional measures promoting incidental learning (e.g Culbertson 2012, 2017, 2023). Plausibility should thus not be manipulated.

The third issue with regards to previous experimental designs lies in the inclusion of complex forms (i.e. biclausal structures). The by-sentence type analysis of Rebuschat (2008) highlights participants' struggle to identify the correct position of the verb. For simple sentences, participants accepted a high proportion of sentences with verbs either in second or final position. This suggests that participants recognised some distributional pattern but failed to form generalisations congruent with the target language. This mirrors the situation observed for L2 learners of German who do not confine finite verbs to the second position in root clauses (Clahsen 1990, Meisel 1997) thereby contrasting L1 learners who place finite and infinite verbs in the correct positions from the earliest stages (Clahsen, Eisenbeiss & Penke 1996). These results stand in contrast to the study of Getz (2018) where participants learnt the V2/V-final contrast well, irrespective of the type and locus of the inflectional morphology. One explanation for the diverging results could be the use of a semi-artificial language in Rebuschat (2008). Learners are known to transfer properties of previously acquired languages to the language that is being learnt (Westergaard 2021b). Consequently, learning a language which

resembles participants' native language very closely would likely cause more transfer effects than a language which exhibits fewer similarities. That is, learners need to overcome non-target consistent transfer of verb placement — a task that is arguably harder when the languages are very similar to each other, thus resulting in a worse performance.[13] The presence of multiple verb positions in the language acquired in Rebuschat (2008) in combination with the presumably strong transfer from learners' native language might cause uncertainties in learners with respect to the correct position of the verb. Irrespective of the underlying processes, the displayed difficulties in placing verbs in the appropriate position entails that if questions regarding V2 are investigated and the V2/V-final distinction does not bear relevance for the research questions, the V2/V-final distinction should be dropped in favour of a simpler design. Such a change should then be concomitant with an increased performance by participants.[14,15] Besides, there are other ways for signalling verb movement. In the Scandinavian V2 languages, the position of the verb relative to a sentence-medial adverb is often used as evidence for movement from the VP/*v*P to a higher position (Vikner 1995, Waldmann 2008, Westergaard 2009b). If ALL did indeed prove to be a suitable methodology for investigating questions related to the diachronic developments of V2, the alternations between V2 and V-final clauses could then of course be reintroduced into the design in future experiments. In §1.5.2, it was argued that the V2/V-final alternation constitutes additional evidence for a V2 grammar in SOV languages like German and Dutch (but crucially not for SVO languages). Hence, scrutinising its effect on learning V2 might be indispensable. However, given that the focus of the present thesis lies on the impact of clause-initial constituents and their distribution, this aspect will not be further explored here.

The fourth and hence final aspect is the way participants' command of the language is measured. As noted in Chapter 1, one of the hallmark features of V2 can be seen in the flexibility of the clause-initial position. That is, no (or at least very few) constraints apply to the grammatical function or category of the clause-initial constituent. The testing items used in Rebuschat (2008) and subsequent work were constructed with novel lexical items but crucially, the acceptability of novel grammatical functions in clause-initial position was not tested. Similarly in Getz (2018), no novel grammatical functions or categories were introduced at test. Arguably, participants need to demonstrate that

[13] Whether transfer can also explain why L2 learners of German struggle with the correct verb placement is beyond the scope of this chapter and has therefore to be left open. See e.g. Schwartz & Sprouse (1996) and Rankin (2014) for a discussion on the role of transfer in L2 acquisition.

[14] Note that this does not entail that participants will learn the verb placement rule perfectly. The results of both Rebuschat (2008) and Getz (2018) point to issues of correctly identifying clause-medial verb positions. While V1 sentences were overwhelmingly rejected, the pattern was much less clear for V3 and V4 sentences. This pattern is not unexpected though. Endress, Scholl & Mehler (2005) found that edges of sequences are more salient to learners than medial parts.

[15] Some support for this conclusion can also be found in Rebuschat (2008). An earlier version of the semi-artificial language learning experiment outlined here included a fourth sentence pattern during training, namely a simple V2 sentence with an analytic verb form (i.e. an auxiliary in second position and a past participle in clause-final position). This sentence type was dropped because too many rules might have a negative impact on learning (Rebuschat 2008: 82).

sentences with e.g. indirect objects or past participles are acceptable to them. Otherwise, the conclusion that participants have learnt a V2 grammar might not be warranted and the language was in fact not sufficiently learnt. That is, future iterations should include a test on how well participants generalise to novel grammatical functions or categories.

The previous discussion has shown that many aspects of previous ALL experiments can be used for investigating V2. However, I have also argued that changes to the design are required, some of which should benefit the learning of participants: (i) uninstructed learning only, (ii) elimination of implausible sentences, (iii) simplified grammar such that no subordinated clauses are introduced and (iv) testing the generalisation of the V2 rule to novel constituent types. The effect of these changes on learners were tested in two experiments which are reported in the following two sections.

2.3 Experiment 1

The preceding discussion has shown that ALL constitutes a suitable methodology for studying V2 phenomena. However, it has also become clear that changes to the previous experiment designs are necessary. Some of these changes should eventually also improve participants' learning of the V2 language. Before the main question of this thesis can be addressed — that is how variability in the input affects learning — it is essential to scrutinise the effect the changes actually have on participants' learning. I therefore conducted an ALL study implementing these changes. The results of previous studies illustrate that, in principle, semi-artificial and fully-artificial languages are viable design choices. The experiment reported here used a semi-artificial language, consisting of an English lexicon and a V2 syntax. Semi-artificial languages have the advantage that participants can focus exclusively on the learning of grammatical structures as the lexicon is already in place. Moreover, the lexical novelty of testing items can be easily controlled. This is particularly useful when measuring the extent to which participants generalise to novel constituent types. Besides, the results of semi-artificial languages have been replicated with fully-artificial languages in non-V2 domains (Culbertson & Adger 2014, Martin et al. 2019, 2020).

The hypothesis under investigation was that V2 can be learnt in an ALL experiment. Although the similarities between Rebuschat (2008) and the current experiment would suggest a statistical comparison, the changes made to the design here prevent me from comparing the two. Furthermore, learning V2 was measured differently, namely in terms of verb placement and generalisation of possible clause-initial elements, in line with the overall goals of this thesis. Consequently, a comparison with Rebuschat (2008) cannot form part of the hypothesis.

2.3.1 Methods

The experiment was preregistered prior to data collection. Ethics approval was granted by the ethics board of the Linguistics and English Language department of The University of Edinburgh (303-1920/1). The experiment was implemented with the JavaScript library jsPsych (de Leeuw, Gilbert & Luchterhandt 2023).

2.3.1.1 Participants

92 participants were tested online after the experiment was advertised to users with US IP addresses on Amazon Mechanical Turk (MTurk). Participants received $4.00 as reimbursement for their participation. Users with at least 1000 approved HITs (Human Intelligence Task) and an approval rate of 98% or higher were eligible for participation. Furthermore, only native speakers of American English with no contact to another language prior to the age of six were asked to participate. Additionally, users were asked not to participate if they had knowledge of a V2 language.[16] Following the application of a pre-registered exclusion criterion (cf. §2.3.1.3), twelve participants had to be excluded from further analysis. One of those participants also indicated knowledge of Afrikaans (a V2 language, cf. §1.2.2) in a post-experiment questionnaire. The final data analysis included 80 participants.

2.3.1.2 Materials

A semi-artificial language was used to construct stimuli sentences. Akin to Rebuschat (2008) and colleagues, English vocabulary was combined with a non-English V2 syntax. All sentences included a subject, a direct object and an adverb. Additionally, sentences comprised either an indirect object or an adjunct (see below). Following the grammatical rules of other V2 languages, the verb appeared always in second position. The adverb was included to signal verb movement from the VP/*v*P to a higher position in the clause and consequently appeared always to the right of the finite verb (Vikner 1995, Waldmann 2008, Westergaard 2009b). Therefore, the adverb will henceforth be referred to as *movement marker*. The grammar of the language licensed subjects, objects and adjuncts in clause-initial position.

For the training phase, 30 unique unordered sets were created, each consisting of a subject, a direct object, an adjunct, a movement marker and a transitive verb. Each constituent was one or two words long. Except for some movement markers, no constituent was repeated across sets. Three different sentences were formed with the words of a set, yielding a subject-initial, an object-initial and an adjunct-initial version. This process is illustrated in (2) for one of the training sets.[17] By balancing the frequency of different clause-initial constituents, the experiment can be used as a baseline for future experiments.

(2) {the musician, the guitar, outside, proudly, plays}
   a. *The musician* **plays** proudly the guitar outside.
   b. The guitar **plays** *the musician* proudly outside.
   c. Outside **plays** *the musician* proudly the guitar.

The lexical items in each set were selected such that the intended meaning of the sentence was readily understandable. As for verbs, the semantics clearly identified

[16] The term *verb second* was not used in the experiment description. Instead, an (almost) exhaustive list of V2 languages was provided.

[17] The annotation as shown in (2) was not visible for participants and is included here for illustrative purposes only.

the agent (subject) and the patient (direct object) irrespective of the order of both. Subjects were consistently animate, whereas direct objects were inanimate. Subjects were realised either as DP or proper nouns, direct objects only as DPs. Adjuncts denoted either a temporal or locative aspect, realised as adverbs, PPs or DPs. The different versions of each set were assigned to one of three blocks. In each block, the frequency of subject-initial, object-initial and adjunct-sentences was balanced. Three different lists were formed from the blocks by Latin Square. For each participant, the order in a block was randomised.

For the testing phase, test items for the judgement task (cf. §2.3.1.3) were created by crossing the factors INITIAL CONSTITUENT and VERB POSITION. The former factor encodes the type of clause-initial constituent which could be either familiar or novel. Direct objects and adjuncts were used as familiar types given their use during training. Subject-initial sentences were not included due to the high similarity with English. Two types of novel constituents were included. The first, complex adjuncts, differed from simple adjuncts in that an additional modifier was included (cf. (3a) and (4a)). The rationale for including more complex constituents was that on the basis of the training items, participants might derive a rule whereby the placement of constituents in clause-initial position is conditioned on the length of a constituent. The second novel clause-initial constituent type were indirect objects (cf. (3b) and (4b)). Importantly, no adjuncts were included in trials with indirect objects.

(3) a. In late April **regrets** *the politician* openly his misconduct.
b. To the coach **throws** *the player* gently the ball.

(4) a. In the workshop the carpenter **saws** rarely a plank.
b. To the co-worker Sylvia **mentions** privately the secret.

The second factor (i.e. VERB POSITION) determined whether the verb appeared in second (3) or third position (4). In all V3 sentences, subjects appeared between the clause-initial constituent and the finite verb.[18] For each factor combination, four sentences were created, yielding a total of 32 test items. Except for some movement markers, no lexical items were repeated from the training items. After the completion of the experiment, a typo in the testing materials was kindly pointed out to me by one of the participants of experiment 2 (cf. §2.4). See Appendix A for the full list of all training and testing materials.

2.3.1.3 Procedure

The experiment was accessible through a web browser and participants completed the experiment online on their personal computer or laptop. At the beginning of the experiment, participants were informed that they would be learning a recently

[18] Structures like these involve the smallest number of derivations; that is non-subjects move to SpecCP, whilst subjects and finite verbs remain in TP, as shown in (i).

(i) $[_{CP}$ The sale$_1$ $[_{C'}$ $[_{TP}$ the administrator$_2$ $[_{T'}$ regulates$_3$ $[_{VP}$ voluntarily in Austria t_2 t_3 $t_1]]]]]$

Figure 2.1: Example trial of training task in experiment 1. Constituents were revealed consecutively, i.e. *the musician*, *plays*, *proudly*, *the guitar* and *outside*. When the sentence became fully visible on the screen, participants were prompted to identify a specific constituent by clicking on it.

discovered dialect of English, which differs from other varieties of English. Specifically, the new dialect would display a greater degree of word order freedom. Participants were also informed that the purpose of the study was to examine how speakers of other English varieties perceive these differences.

The experiment was divided into two parts — a training phase and a testing phase. Participants were assigned randomly to one of the three training lists. Before commencing with the actual training, participants were briefed that sentences would be displayed on the screen for which they had to identify either the actor (i.e. subject), the passively involved entity (i.e. object), the action (i.e. verb) or either the time or place of the action (i.e. temporal or locative adjunct). These lay terms were used instead of linguistic terminology to avoid possible confusion on the side of the participants. On each trial, the stimulus sentences was revealed constituent by constituent, with a delay of 500ms between constituents. Once the sentence became fully visible on the screen, participants were prompted to identify one of the aforementioned constituents, see Figure 2.1. Each of the four constituents were enquired 18 times whereby temporal and locative adjuncts were treated as one constituent type.[19] Feedback was provided after each trial: The correct constituent appeared in green font and, if participants answered incorrectly, the wrong answer was shown in red in addition to the correct answer. To disincentivise participants from clicking through the training without properly interacting with the stimuli, feedback screens were displayed for 2000ms in case of an incorrect answer. If participants selected the correct constituent, the feedback was visible for 450ms before the next trial started. Participants whose accuracy fell below 90% during training were excluded from the analysis. Participants completed 90 training trials.

After completing the training, acceptability judgements were elicited from participants. Participants had to decide whether the sentence presented in any of the 32 trials would be uttered by a speaker of the dialect. Participants had to select either ‘yes’ or ‘no’. No feedback was provided. An example trial is given in Figure 2.2. In the final part of the experiment participants filled in a short questionnaire enquiring their language background, their assumptions about the purpose of the experiment and whether they had developed any strategies to complete the experiment.

[19]The number of trials where participants had to select a temporal or locative adjunct was balanced.

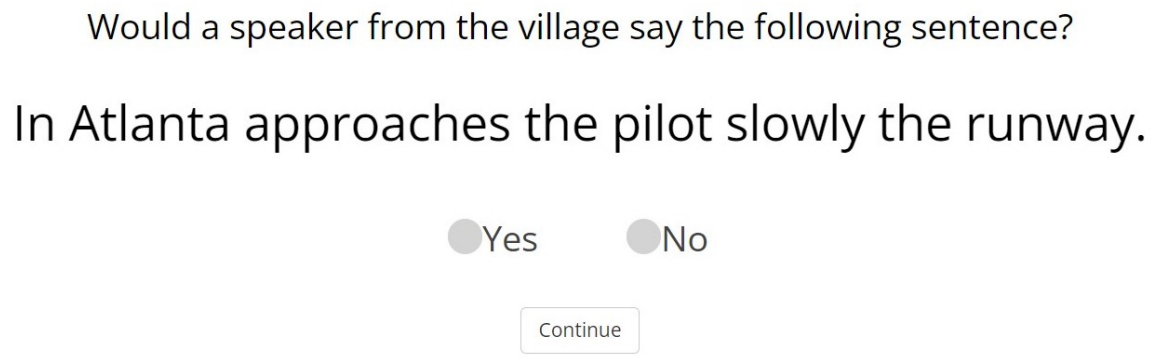


Figure 2.2: Example trial of testing task in experiment 1. Participants had to determine whether the provided sentence could be produced by a speaker of the language.

2.3.2 Predictions

As outlined earlier, the hypothesis investigated in the experiment was that V2 can be learnt in an ALL experiment. Learning V2 was operationalised in terms of the defining features of V2: verb position and XP-fronting (cf. Chapter 1). That is, participants need to learn that verbs are confined to the second position in the clause. Additionally, participants need to extrapolate that any constituent type (other than the finite verb) can be placed in the clause-initial position. For the present experiment, this means complex adjuncts and indirect objects.[20] Three predictions could be derived from the hypothesis in conjunction with this operationalisation. First, participants should rate V2 structures that they have seen during training above chance level. This prediction acts as a sanity check for the experiment. Second, participants should generalise this pattern to structures that they have not previously seen. Finally, participants should be able to discriminate between unfamiliar V2 structures and unfamiliar non-V2 orders whereby the former are rated higher than the latter.

2.3.3 Results

The data gathered in experiment 1 and experiment 2 were analysed in R (R Core Team 2022) using ggplot2 for plotting (Wickham 2016) and lme4 for statistical analysis (Bates et al. 2015). Wald tests were used to obtain *p*-values of the model coefficients. The standard alpha level of 0.05 was used to determine significance.

[20] Note that according to this operationalisation the grammatical function (e.g. subject, object) and not the grammatical category (e.g. DP, PP) of the clause-initial constituent is the determining factor in the grammar. In principle, grammatical rules could make reference to either of them (cf. Yang (2000, 2002) and Lightfoot (1999, 2006)). For the present experiment, grammatical functions were used due to their usage in describing grammars (e.g. SVO, SOV). As mentioned above, the results presented in Chapters 3 and 4 will prompt a reformulation of the hypothesis in that variability in both grammatical functions and categories is assumed to benefit the learning of a V2 language.

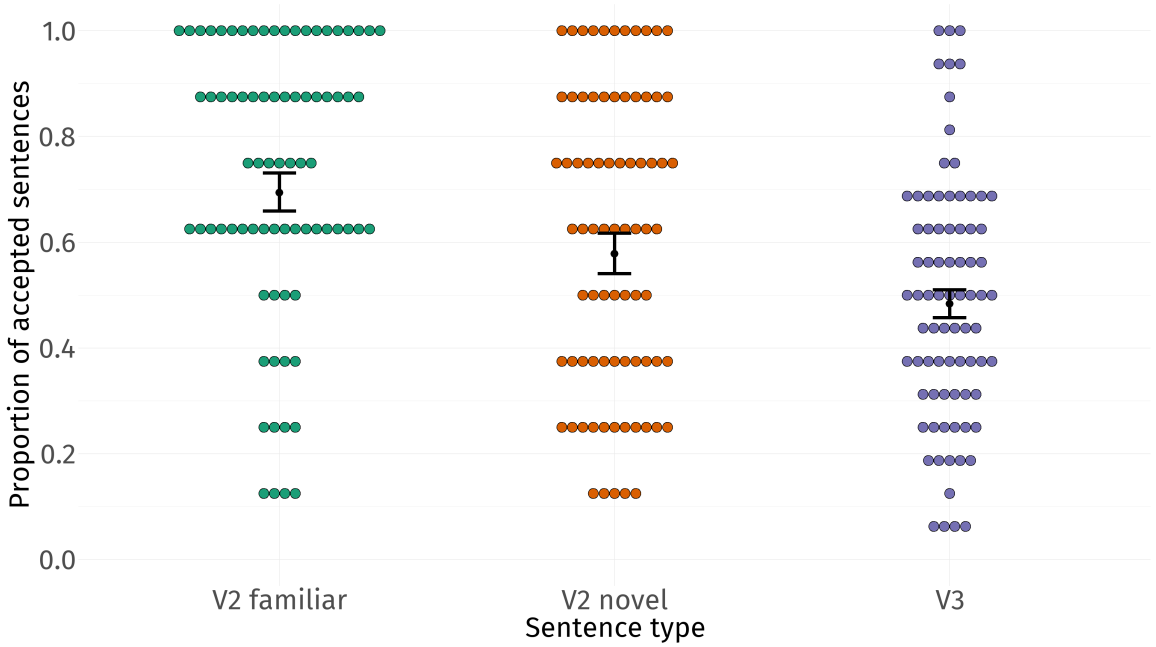


Figure 2.3: Acceptance rates of sentences of the types V2-familiar, V2-novel and V3 in experiment 1. V2-familiar sentences comprises all V2 sentences with clause-initial direct objects and simple adjuncts. Similarly, V2-novel sentences encompass V2 sentences with initial indirect objects and complex adjuncts. All sentences with V3 order are subsumed under the last type that is V3. Coloured dots represent participants' ratings, black dots the means. Error bars indicate bootstrapped 95% confidence intervals of the mean. All three predictions were confirmed.

2.3.3.1 Hypothesis-confirming analysis

To assess the predictions outlined in §2.3.2, the two factors INITIAL CONSTITUENT and VERB POSITION were used to create a new factor called SENTENCE TYPE. The V2 sentences with initial direct objects and simple adjuncts were subsumed under the label *V2-familiar*, whereas V2 sentences with initial indirect objects and complex adjuncts were conflated as *V2-novel*. All V3 sentences were grouped together as *V3*. Participants' ratings for the three sentence types are depicted in Figure 2.3. To test the first prediction — that is whether the ratings for V2-familiar sentences are above chance — I fitted an intercept-only mixed-effect logistic regression model to the judgement data from V2-familiar sentences. The model included by-participant and by-item random intercepts. The model revealed that the first prediction was indeed borne out as the intercept was significantly above chance ($\beta = 1.19$, SE $= .23$, $p = 3.73 \times 10^{-7}$).

The second and third prediction were assessed in a further model. I fitted a mixed-effect logistic regression model to the V2-novel and V3 data. The model included SENTENCE TYPE as fixed effect. The model also included by-participant and by-item

random intercepts as well as by-participant random slopes for SENTENCE TYPE. The fixed effect was treatment coded with V2-novel as reference level. The second prediction (i.e. acceptance rates of V2-novel sentences should be above chance) was assessed by examining the intercept of the model. The model showed that the intercept is significantly above chance, thus again confirming my prediction ($\beta = .45$, SE $= .20$, $p = 0.02$). For the third prediction, i.e. the discrimination of novel but grammatical V2 sentences and ungrammatical structures, I examined the simple effect of SENTENCE TYPE. As can be seen in Figure 2.3, the acceptance rate of V2-novel sentences is generally higher than the one of V3 sentences. This was also confirmed by the model ($\beta = -0.52$, SE $= .22$, $p = .02$).

2.3.3.2 Exploratory analysis

The hypothesis did not warrant any conclusions about the relation between V2-familiar and V2-novel sentences. The contrast, however, does nonetheless bear relevance for the goals of this thesis. If the ratings for V2-familiar and V2-novel did not differ significantly, this might provide valuable insights into the readiness with which participants extrapolate XP-fronting to novel constituent types. This question was thus addressed in a further logistic mixed-effect model. The model was fitted to the V2-familiar and V2-novel data only. The model included SENTENCE TYPE as fixed effect as well as by-participant and by-item random intercepts. Furthermore, the model included random by-participant slopes for SENTENCE TYPE. SENTENCE TYPE was treatment coded with V2-familiar as baseline. Participants assigned significantly lower ratings to V2-novel sentences than to V2-familiar sentences ($\beta = -0.74$, SE $= .23$, $p = .002$).

I also scrutinised the responses participants provided in the post-test questionnaire. Specifically, I was interested if participants developed any strategies that would hint at the identification of syntactic rules underpinning the language. Two participants noted the frequent inversion of subjects and finite verbs — a characteristic feature of the employed language. Albeit correlating with the performance in the judgement task of one participant (i.e. higher ratings for V2-familiar as well as V2-novel sentences but low ratings for V3 sentences), the other participant showed a less clear discrimination between grammatical and ungrammatical sentences. Two further participants stated that they had paid attention to the position of certain elements in the sentences such as the subject, but crucially this was reflected only in a good performance by one participant. Other reported strategies were less elucidating. For instance, one participant reported reading all sentences backwards. Although this might indicate that sentences were challenging for at least some participants, nothing else can be gained.[21]

2.3.4 Discussion

The goal of this chapter is to develop an experimental design that allows studying V2 with ALL. After discussing a series of studies that have used artificial V2 languages,

[21] This participant was not excluded form the data analysis because employed strategies were not defined as exclusion criterion in the preregistration. Besides, the acceptability ratings of this participant lied within the range of other ratings.

four modifications to previous designs were distilled: (i) no explicit instruction of grammatical rules, (ii) elimination of trials with implausible sentences, (iii) usage of a simplified language without multiple positions for verbs and (iv) inclusion of tests for measuring how well participants generalise XP-fronting to novel constituent types. Whilst (iv) is a new measure of learning and should as such not have an effect on learning, (ii) and (iii) should enhance participants' learning performance compared to previous studies, in particular Rebuschat (2008). (i), in turn, should ensure that the influence of learning is measured and not some other cognitive processes. The purpose of the present experiment was therefore to investigate how these changes affect the learning performance of participants. Although some design choices would suggest a direct (statistical) comparison with Rebuschat (2008), the design changes render this impossible. As a result, I tested a simplified hypothesis. Specifically, I hypothesised that V2 can be learnt in an ALL experiment. Learning a V2 language was defined in terms of the acquisition of the correct verb placement as well as the extrapolation of XP-fronting to the clause-initial position to constituents that have not been previously seen. Participants were taught a semi-artificial language comprising an English vocabulary but a V2 syntax. Participants were then asked to judge whether sentences with either familiar or unfamiliar constituent types in clause-initial position were grammatical. For familiar constituents, I used direct objects and adjuncts. Unfamiliar constituents, on the other hand, were represented by indirect objects and complex adjuncts — that is adjuncts that were longer than those seen during training.

The analysis of the testing data revealed that participants were able to learn the structures they were trained on. This was reflected in a high acceptance rate of V2-familiar sentences. To assess whether participants learnt a genuine V2 grammar or a grammar which merely licences subject-initial, object-initial and adjunct-initial sentences, participants' ratings for V2 sentences with novel constituent types were scrutinised in more detail. The results show that participants generalised XP-fronting to novel constituent types. Crucially, the acceptance rate of V2-novel sentences was higher than the one for ungrammatical V3 sentences. This suggests that participants indeed learnt a V2 grammar and generalised the pattern to novel constituent types. My exploratory analysis suggests, however, that participants were more hesitant with generalising XP-movement than accepting familiar V2 structures. Yet the fact that all three predictions were borne out still provides compelling evidence for the learnability of an artificial V2 language even after implementing the design changes.

To summarise, as far as the hypothesis is concerned, the changes suggested in §2.2.3 to earlier experimental designs and implemented here were successful in that the V2 language was learnt. Recall that I also predicted improved learning by participants. Although a statistical comparison with earlier work is not possible, the results can still be compared descriptively. Rebuschat (2008) will serve as the main point of comparison due to the overall large similarities to the current experiment and sufficiently detailed reported results. One of the sentence types used for testing in Rebuschat (2008) are V2 sentences. This type is partially comparable to V2-familiar sentences in the present experiment. The reported acceptance rate of 73.4% (cf. Table 2.2) and the one observed here (69.4%) are similar. A contrast between the two experiments arises in the ratings of

V3 sentences though. In my experiment, the acceptance of V3 sentences is substantially higher (48.4%; 22.9% in Rebuschat 2008).[22] These findings could be interpreted as a low efficacy of the introduced changes. There are, however, two arguments against this interpretation. First, participants received less training in the current experiment than in Rebuschat (2008) — in fact, 30 fewer training trials were included. In the light of the comparable ratings for V2 sentences across both experiments, additional evidence for the confinement of the verb to the second position would presumably not increase the acceptance rate for V2 sentences (though see discussion below) but lead to a higher rejection rate of V3 sentences. That is, participants would be more confident that verbs are *not* realised in the third position of the clause.

The second argument centres on the training and testing materials. More precisely, the materials used in Rebuschat (2008) exhibited a greater similarity to participants' native language (i.e. English). While in the current experiment subjects, adjuncts and direct objects were attested clause-initially in the training materials, only subjects and adjuncts occurred in the clause-initial position in Rebuschat (2008). English speakers are exposed to a large proportion of subject-initial sentences in their environment due to English being a SVO language. In contexts where non-subjects occupy the clause-initial position, adjuncts account for the majority of all cases. Doherty (2005: 192), for instance, observed that topicalised adjuncts (5a) are significantly more frequent than topicalised objects (5b).[23]

(5) a. In the dark, *the mouse* **gulped down** a spritz.
b. A slice of mature cheddar, *the mouse* **enjoys** occasionally.

A similar picture arises in contexts that resemble V2 structures more closely. Negative inversion — one of the residual V2 contexts in English — has been described as a construction where negated adjuncts are being fronted (6a) (Rudanko 1982), even though object-initial sentences (6b) are attested, as Holmberg (2015: 344) notes.

(6) a. Under no circumstances **will** *I* believe your alternative facts.
(Sailor 2020: 126)
b. None of them **would** *I* recommend.
(Holmberg 2015: 344)

Similarly in cases of locative inversion, the verb is typically preceded by a PP (7a), albeit adverbs (7b) can occur clause-initially as well (Green 1980: 582, Birner 1995: 240, Postal 2004: 16).[24] Quotative inversion constitutes the only non-negative context in which objects can be preposed. In this type of inversion, the verb follows its direct speech complements or parts thereof (Collins & Branigan 1997, Bruening 2016). An example for quotative inversion is provided in (8).

[22] In this regard, the results of my experiment align with those of Getz (2018).
[23] Structures like (5) are generally rare. Yang (2000: 242) found that less than 10% of all sentences in the Penn Treebank are topicalisations.
[24] Note that the term *locative* is a misnomer given the existence of sentences like (7a) where the clause-initial PP does not denote a locative aspect (Postal 2004: 17, Bruening 2010: 46).

(7) a. Against that proposal **can** be objected *the fact* [that no one is entirely logical].
(Postal 2004: 17)
b. Outside **stood** *a little angel*.
(Green 1980: 582)

(8) "Where to?" **asked** *the driver* of his passenger.
(Collins & Branigan 1997: 4)

The different structures illustrate that English speakers are generally more familiar with clause-initial adjuncts than with clause-initial (direct) objects. Coming back to the comparison with Rebuschat (2008), the inclusion of object-initial sentences in training thus rendered the language in the present experiment more dissimilar to participants' native language. The language in Rebuschat (2008), on the other hand, exhibited greater similarities to participants' native language. The closer typological proximity of Rebuschat's (2008) language to English could then explain the better learning. This effect might have been further consolidated by the testing materials. V2-familiar sentences featured only direct objects and adjuncts in clause-initial position. The corresponding V2 testing sentences of Rebuschat (2008) were exclusively adjunct-initial. That is, high acceptance rates were more likely to occur. Consequently, achieving a similar acceptance rate to Rebuschat (2008) in the current experiments despite the more difficult language provides a compelling argument for the efficacy of the suggested changes.[25,26]

Despite the demonstrated effectiveness of the implemented changes compared to previous experimental designs, it would still be desirable to improve participants' learning further. Investigating V2 experimentally requires participants to have a good command of the grammar. Hence, one should strive to maximise the learning performance of participants. I will therefore explore possible ways of fostering the learning in the remainder of this section. A striking difference between the semi-artificial language studies and the present experiment is the way participants are trained. Recall that participants in Rebuschat (2008) had to repeat each training sentence after presentation. This part of the procedure was not implemented in the current experiment even though the inclusion of production trials or some other production component is a common practice in ALL experiments, both with children and adults. The production could either be in the form of oral productions (e.g. Culbertson & Newport 2017, Tal & Arnon 2022) or written productions (i.e. typing text) (Smith & Wonnacott 2010). In fact, evidence for the benefits of production during training has been provided by Hopman & MacDonald (2018). In their study, two groups of participants were compared who received training in an artificial language in the form of passive exposure coupled either with comprehension trials or production trials. Hopman & MacDonald (2018) found that participants in the production condition exhibited significantly faster reaction times across the board. In addition, participants in the production condition were more accurate when tested on agreement suffixes. However, participants in the production

[25]For a general discussion on how participants' L1 affects their learning performance, see Chapter 3.
[26]This explanation leaves open why V3 sentences in particular receive lower ratings.

condition and comprehension condition were equally accurate when judging whether the word order of a sentence was correct. At first glance, this might suggest that the inclusion of production trials in a V2 experiment will not improve participants learning given the focus on a word order phenomenon in the present study. The difference in reaction times contradicts such a conclusion though. The lower reaction time indicates that representational differences exist in that participants in the production condition dispose of a more fully-fledged and/or better accessible representation. Additional evidence for the beneficial role of production during learning can also be found in Zimmerman et al. (2009). They show that the number of conversational turns in child-adult interactions are positively correlated with the language development of children. In view of the outlined findings, including production trials in the experiment should foster participants' learning.

Another difference between Rebuschat (2008) and the current experiment lies in the testing modality. While Rebuschat (2008) tested participants in the lab, I tested participants online. It is conceivable that this might have affected the results for two reasons (cf. Uittenhove, Jeanneret & Vergauwe 2023). On the one hand, one does not tap into the homogeneous population typically used for behavioural research, i.e. undergraduate university students. This disparity might have been further exacerbated by the COVID-19 pandemic when many people worked from home or were unemployed and thus forced to seek alternative income streams. On the other hand, participation does not take place in an environment controlled by the experimenter. Consequently, participants might be less focused (e.g. due to ambient noise) which could have detrimental effects on language learning. Investigations into the comparability of crowdsourcing and lab-based experimenting have been conducted, inter alia, for cognitive psychology (e.g. Crump, McDonnell & Gureckis 2013) and linguistics (Schnoebelen & Kuperman 2010, Sprouse 2011) with authors generally being able to reproduce findings from the lab in an online environment. More recent work, however, has cast doubt on those findings. Peer et al. (2021) and Uittenhove, Jeanneret & Vergauwe (2023) who compare MTurk with other crowdsourcing platforms discover the data quality of MTurk to be inferior to other platforms, in particular Prolific.[27] This might be attributable to a lower motivation and/or lower attention (e.g. due to distraction) to the experimental task of (recent) MTurk participants. Moreover, MTurk participants might be more likely non-human (i.e. a bot). Uittenhove, Jeanneret & Vergauwe (2023) further compare the results to lab-tested and online-tested university students. Interestingly, they discovered strong similarities between the results of online-tested university students and Prolific users. Data gathered from lab-tested students in turn exhibited only a minor advantage over the data from web-tested students and Prolific users.[28] Two conclusions can be drawn from this. First, using MTurk could have contributed to the relatively low per-

[27] This does not mean that earlier research should be discarded. Arechar & Rand (2021) observed changes in the composition of the MTurk participant pool following the COVID-19 pandemic, *pace* Moss et al. (2020). The changes in the MTurk population might thus explain the observed divergence from earlier work.

[28] Deeming lab-tested participants as gold standard of experimentation is not unproblematic. Many of these studies sample from western, educated, industrialised, rich, democratic (WEIRD) populations which may not be as representative as previously claimed (Henrich, Heine & Norenzayan 2010).

formance. Second, Prolific offers itself as a viable alternative for online data collection. Participants recruited from Prolific could perform better at the learning task.

In this section, I have presented the results of an artificial language learning experiment. The results strongly suggest that V2 can be learnt in a short artificial language learning experiment. In the discussion, it was argued that participants performed better at the learning task compared to early studies albeit somewhat obscured by a more complex language. I subsequently argued that learning can still be further improved by (i) recruiting participants from Prolific instead of MTurk and by (ii) including production trials during training. In the following section, an experiment is presented that implemented these changes.

2.4 Experiment 2

Experiment 1 has demonstrated that V2 can be learnt in a semi-artificial language learning experiment. The changes I made to the experiment design compared to Rebuschat (2008) have been effective and fostered learning, yet a better learning outcome is still desirable for further investigations. In the subsequent discussion, I identified two modifications that should further improve participants' learning. First, production during learning has been shown to facilitate learning (Hopman & MacDonald 2018). Including production trials in the training should thus be beneficial. Second, recent evidence suggests that the tested population (i.e. MTurk users) is less reliable and performs significantly worse in behavioural experiments than participants tested in the lab or those recruited from other crowdsourcing platforms such as Prolific. I thus reran the experiment with production trials during training as well as testing with participants recruited from Prolific. The hypothesis remained unaltered.

2.4.1 Methods

Experiment 2 was preregistered before data collection commenced. The experiment received ethics approval from the ethics board of the Linguistics and English Language department of The University of Edinburgh (180-2021/2). The experiment was again implemented with the JavaScript library jsPsych (de Leeuw, Gilbert & Luchterhandt 2023).

2.4.1.1 Participants

94 participants were recruited on Prolific. Each participant received £4.36 as reimbursement for their participation. In-built Prolific filters were used to restrict participation to participants who are US nationals, who are English monolinguals and who had been raised monolingually. Additionally, participants whose subject at university did involve English literature, English language or languages more broadly were excluded from participation. Finally, only participants whose approval ratings were at least 95% were

Hence, the results should be cautiously interpreted as it remains a possibility that sampling from a less WEIRD participant pool is the driving factor for the diverging results.

able to participate. Based on the pre-registered exclusion criterion, 19 participants were excluded due to low performance during training (cf. §2.4.1.3) and one participant had to be excluded as they indicated knowledge of a V2 language in a post-test questionnaire. 74 participants could therefore be included in the analysis. Although some participants indicated note-taking during the experiment, no further exclusions were made as this was not preregistered as exclusion criterion. Furthermore, the post-test questionnaire did not enquire whether notes had been taken during the experiment.

2.4.1.2 Materials

The training and testing materials used in experiment 2 were identical to those in experiment 1. This meant that the materials also contained the typo mentioned in experiment 1. Due to an error in the experiment code, merely two (instead of three) training lists were constructed.[29] The only difference between the two experiments in terms of the materials was the addition of items for production testing. These testing items consisted of unordered sets of constituents. Three different set types were constructed: the first type (9a) comprised only constituent types participants have seen during training (i.e. subject, direct object, adjunct and movement marker). In the other two sets the adjunct was replaced with a novel constituent type, either a *complex adjunct* with an additional modifier compared to simple adjuncts (9b) or an indirect object (9c). The order in those sets was randomised for each participant. The same design principles as in experiment 1 were applied to the construction of the items.

(9) a. {the general, declares, hastily, victory in Waterloo}
b. {Brianna, refutes, energetically, the rumour, in late January}
c. {the investigator, submits, belatedly, the report, to the prosecutor}

2.4.1.3 Procedure

The start of the experiment was identical for participants in both experiments. That is, participants accessed the experiment on their personal computer or laptop and they were informed that they would be learning a recently discovered dialect of English. The experiment comprised a training phase and a testing phase. The training phase was divided into two parts: sentence reading trials and production trials. The reading trials were identical to those from experiment 1 apart from two differences. First, the number of reading trials was reduced to 30 (instead of 90). This corresponds to the design of Hopman & MacDonald (2018) where the number of production trials exceeded the number of passive exposure trials. Second, the feedback screen for incorrectly identified constituents was displayed for a shorter period, namely 850ms. Participants' performance on reading trials was used as exclusion criterion. Whilst I maintained the 90% threshold from experiment 1, the actual permissible number of erroneous choices was limited to 3 or less due to the reduced number of trials.

The reading trials were followed by 60 production trials. At the beginning of each trial, the initial constituent of a sentence was provided together with four blank lines,

[29] Note that this error did not affect experiment 1.

Figure 2.4: Example trial for production task during training in experiment 2. At the beginning of each trial, only the initial constituent was given, whereas the remaining constituents were shown underneath. To construct sentences, participants had to click on the words which where then added to the sentence. If participants were not satisfied with their production before submitting their answer, they could reset the sentence and start afresh. Participants had to use all of the provided constituents.

as illustrated in Figure 2.4. The four remaining constituents were shown as clickable buttons below the initial constituent and the four lines. Participants were instructed to complete the sentence with all of the provided constituents. If not all constituents were used, a message was displayed reminding participants to use all constituents. Participants could reset their answer before submitting it. After each trial, feedback was shown to participants. If constituents were placed in the incorrect position, they were highlighted in red. In addition, the correct sentence was displayed below participants' answer. Feedback was shown for 1500ms if the produced sentence was correct and 3000ms if incorrect. The objective of long feedback times was to disincentivise participants from merely randomly clicking on buttons.

After completing the training, participants proceeded to the testing phase of the experiment. The testing also comprised two parts. In the first part, participants were asked to produce sentences. The procedure was identical to the one seen during training with the crucial difference that no initial constituent and no feedback was provided. A total of 12 trials were completed by participants. The order of the trials was randomised for each participant. The production trials were followed by 32 judgement trials. The procedure was identical to the one in experiment 1. The experiment concluded with the same post-test questionnaire given to participants in experiment 1.

2.4.2 Predictions

The hypothesis investigated in experiment 2 remained unaltered from experiment 1. That is, it was hypothesised that a V2 language can be learnt in an ALL experiment.

Learning a V2 language was defined as extrapolation of the verb placement rule and the flexibility of the clause-initial constituent to novel structures. From this, I derived testable predictions for the obtained production data and the judgement data. As for the former, I predicted that (i) frequency with which finite verbs are placed in second position is above chance independent of the familiarity of the sentence type. Additionally, (ii) participants should produce V2 sentences with clause-initial unfamiliar constituents at higher than chance level. With regards to the judgement data, the same predictions as in experiment 1 were made (cf. §2.3.2): (i) participants should rate V2 structures they have been familiarised with during training at higher than chance levels. Furthermore, (ii) participants should rate unfamiliar V2 structures above chance if they have learnt V2. The final prediction for the judgement data stated that (iii) participants should be able to discriminate between unfamiliar V2 structures and unfamiliar V3 structures whereby the ratings for the former should be higher than those for the latter.

2.4.3 Results

2.4.3.1 Hypothesis-confirming analysis

Figure 2.5 illustrates the proportion of produced V2 sentences by learners broken down for different sentence types (simple adjunct, complex adjunct, indirect object). The label for each category refers to the additional constituent present apart from subjects, direct objects, verbs and movement markers. As aforementioned, complex adjuncts and indirect objects constitute novel types of constituents while participants are familiar with simple adjuncts. I first test whether participants place verbs in second position above chance level irrespective of the sentence type. I fitted a mixed-effects logistic regression model to all produced sentences. The dependent variable was verb position (V2 = 1, non-V2 = 0). As defined in the preregistration, the model included SENTENCE TYPE as fixed effect, and by-participant and by-item as random intercepts and by-participant random slopes. SENTENCE TYPE was treatment coded with simple adjunct as baseline. Due to singular fit, the by-item random intercepts and by-participant random slopes were dropped. The model indicated that participants produced V2 orders in simple adjunct sentences above chance ($\beta = 3.98$, SE $= .67$, $p = 1.49\times^{-15}$). The model further showed that participants produced significantly fewer V2 orders in complex adjunct sentences ($\beta = -0.69$, SE $= .30$, $p = .02$) and indirect object sentences ($\beta = -0.61$, SE $= .30$, $p = .04$). That is my first prediction was only partly confirmed: Even though verbs were produced in second position above chance level, the frequency of V2 sentences is significantly lower in sentences with novel constituent types.

My second prediction concerned the type of clause-initial constituent in participants' productions.[30] I tested whether participants placed unfamiliar constituent types in clause-initial position in V2 sentences at higher than chance level. Figure 2.6 shows the proportion of familiar and novel constituent types in V2 sentences. To assess this prediction, I fitted a mixed-effects logistic regression model to all V2 sentences with a novel constituent type from the production trials. Each sentence was coded for whether a novel constituent (= 1) or familiar constituent (= 0) was placed in the clause-initial

[30]For a breakdown of the different clause-initial constituents, see §3.2.

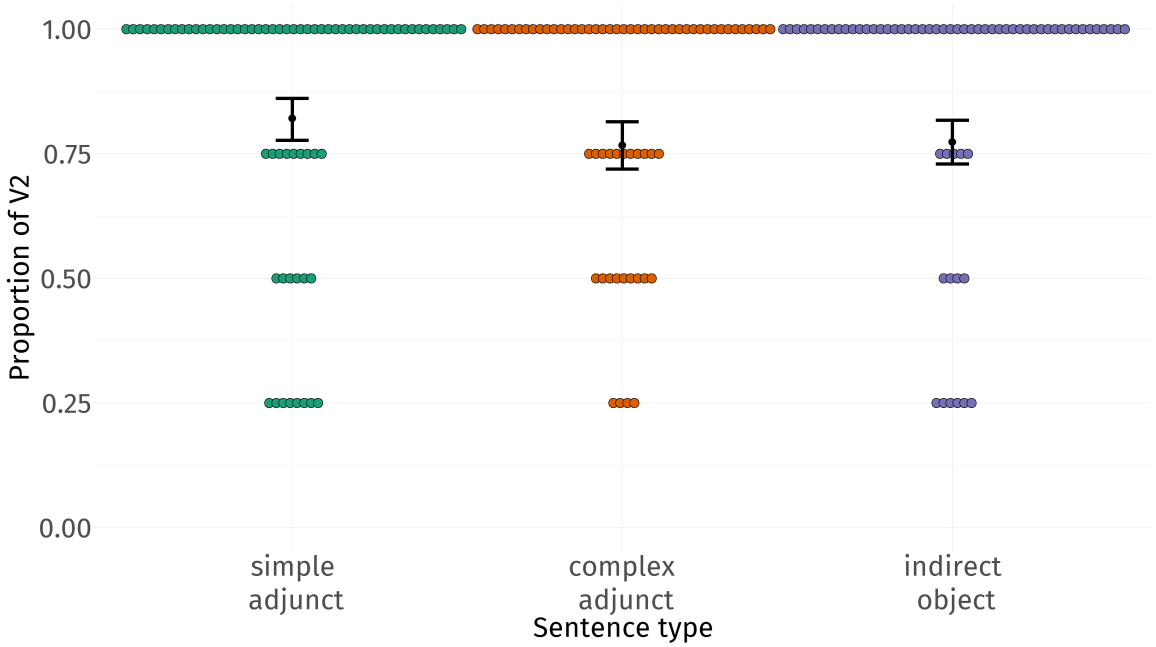


Figure 2.5: Proportion of produced V2 sentences by sentence type in the production test of experiment 2. Sentence type indicates which additional constituent type was provided to participants apart from subjects, verbs, direct objects and the movement marker. That is, the sentence type does not describe the type of clause-initial constituent. Error bars indicate bootstrapped 95% confidence intervals. Coloured dots represent the proportion of participants' V2 production, black dots the mean. Participants placed verbs significantly more frequently in second position in sentences with simple adjuncts than in sentences with complex adjuncts or indirect objects.

position. The model was an intercept only model with by-participant random intercepts. The model revealed that participants realised novel constituent types in clause-initial position significantly below chance ($\beta = -2.47$, SE $= 0.31$, $p = 9.9 \times 10^{-6}$). My prediction was therefore not confirmed.[31]

Turning to the judgement data, I followed the same procedure for the analysis as in experiment 1. In a first step, a new factor SENTENCE TYPE was created by combining different levels of INITIAL CONSTITUENT and VERB POSITION. V2 sentences with direct objects and simple adjuncts in clause-initial position are summarised as *V2-familiar*, while V2 sentences with clause-initial indirect object and complex adjuncts are subsumed under *V2-novel*. All V3 sentences, irrespective of their clause-initial constituent, are grouped as *V3*. The acceptance rates for all three sentence types are visualised in Figure 2.7. To assess my first prediction (i.e. ratings for V2-familiar sentence should be above

[31] Remarkably, some participants also placed the movement marker in the clause-initial position in their productions.

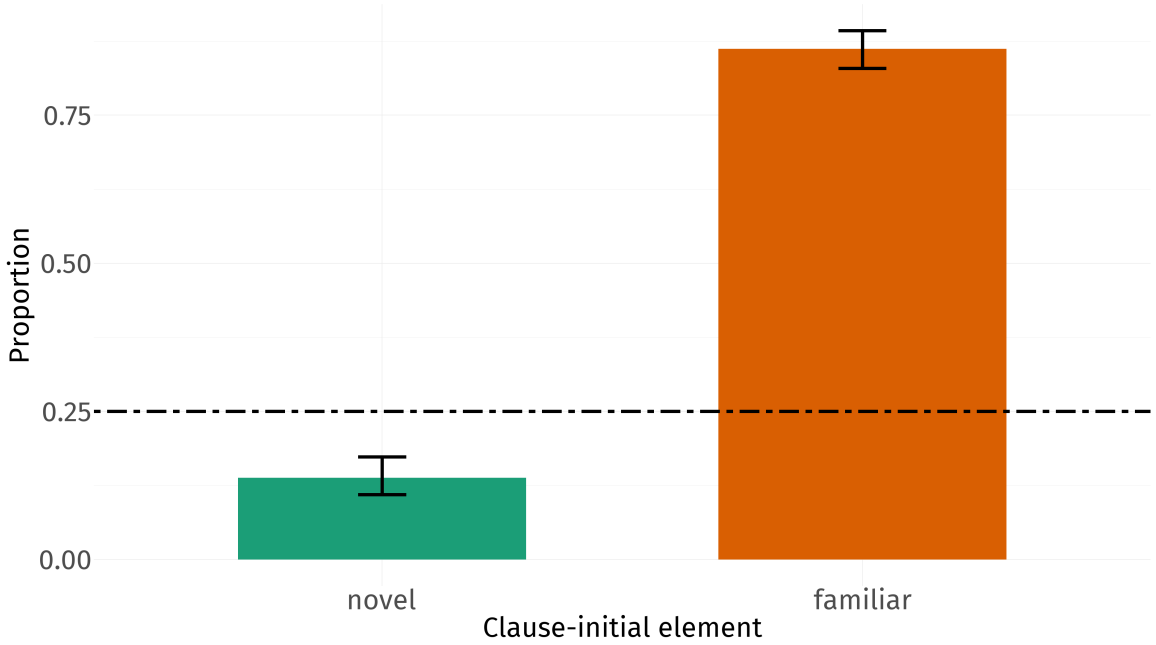


Figure 2.6: Proportion of familiar and novel constituent types in produced V2 sentences in experiment 2. Both constituent types are the complement to each other. Consequently, the values will add up to 1. Error bars indicate bootstrapped 95% confidence intervals. The dotted line represents chance level. Participants did produce V2 sentences with novel constituents in clause-initial position significantly below chance. The prediction was thus not confirmed.

chance level), a mixed-effects logistic regression model was fitted to the V2-familiar data. The model was an intercept-only model and included by-participant and by-item random intercepts. The model showed that participants rated V2-familiar sentences significantly above chance level ($\beta = 1.66$, SE $= .30$, $p = 1.78 \times 10^{-8}$).

The second and third prediction for the judgement data (i.e. V2-novel sentences are rated above chance while also being higher rated than V3 sentences) were addressed in a further model. I fitted a mixed-effects logistic regression model to the V2-novel and V3 data. The model included SENTENCE TYPE as fixed effect. Additionally, the model was fitted with by-participant and by-item random intercepts as well as by-participant random slopes for SENTENCE TYPE. SENTENCE TYPE was treatment coded with V2-novel as reference level. To test the second prediction, I scrutinised the intercept. The model confirmed my prediction in showing that the acceptance rate of V2-novel sentences is significantly above chance ($\beta = .87$, SE $= .26$, $p = 7.39 \times 10^{-4}$).

I assessed the third prediction by investigating the simple effect of SENTENCE TYPE. The model indicated that my third prediction was again confirmed. The acceptance rate for V3 sentences are significantly lower than V2-novel sentences ($\beta = -2.33$, SE $= .36$, $p = 1.11 \times 10^{-10}$). In summary, almost all of my predictions were confirmed in

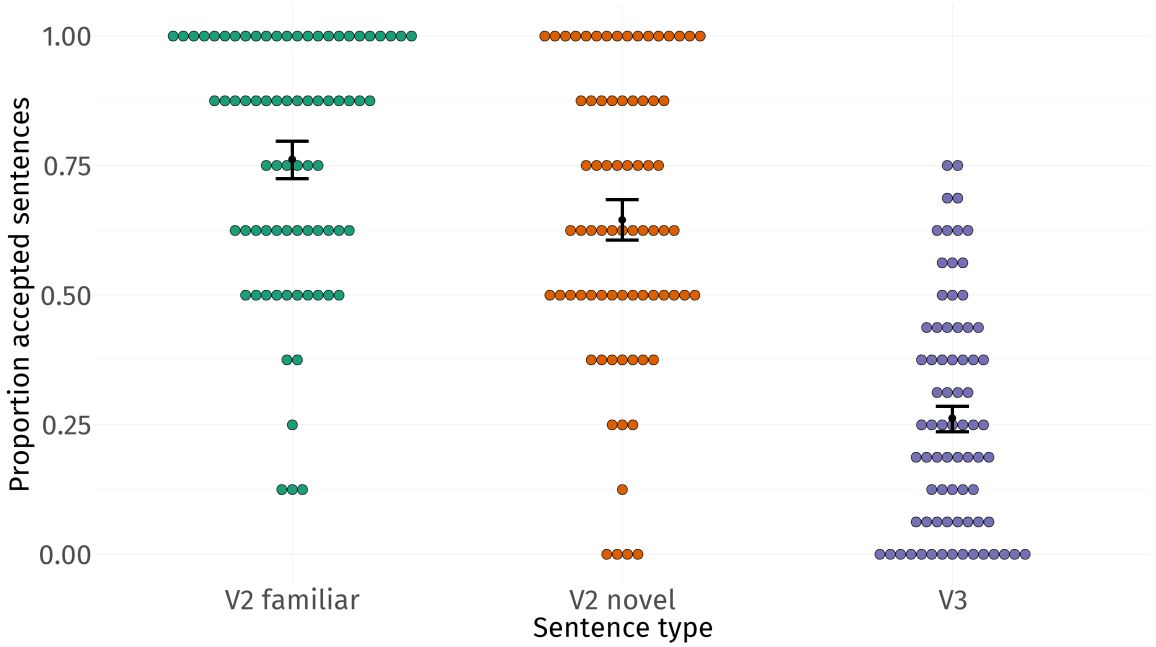


Figure 2.7: Acceptance rates of V2-familiar, V2-novel and V3 sentences in experiment 2. V2-familiar is the category of V2 sentences with clause-initial direct objects and simple adjuncts. V2-novel sentences comprise V2 sentences with unfamiliar constituent types in clause-initial position, that is indirect objects and complex adjuncts. V3 sentences subsumes all sentences exhibiting a V3 order. Coloured dots represent participants' mean ratings, black dots the mean of the means. Error bars represent bootstrapped 95% confidence intervals of the mean. All three predictions were confirmed.

experiment 2. The only prediction that could not be confirmed was the frequency with which participants place novel constituent types in the clause-initial position in their productions.

2.4.3.2 Exploratory analysis

In the exploratory analysis of experiment 1 (§2.3.3.2), V2-familiar and V2-novel sentences were compared. The motivation for this particular comparison was to assess the degree to which participants extrapolated V2 to novel structures. If no significant difference was observed, it would suggest that participants generalise readily without issues. I repeated the same analysis for experiment 2. A mixed-effects logistic regression model was fitted to all V2-familiar and V2-novel judgements. The model included SENTENCE TYPE as fixed effect. Additionally, by-participant and by-item random intercepts were fitted as well as by-participant random slopes for SENTENCE TYPE. The fixed effect was treatment coded with V2-familiar as baseline. The model revealed that the acceptance rate of V2-novel sentences was significantly lower than the one of V2-familiar sentences

($\beta = -0.80$, SE = .30, $p = 0.007$). That is, the same pattern as in experiment 1 was observed.

Movement markers (i.e. sentence-medial adverbs) were included to signal to participants that verbs are raised from VP/*v*P to a higher position in the clause. Verbal movement has been argued to be one of the core properties of V2 (cf. §1.2.3). It is therefore crucial that participants have recognised the function of the movement marker. Otherwise, objections could be raised as to the V2 nature of the acquired grammar. In particular, one could argue that the grammar might merely resemble a V2 grammar on the surface. Although no direct test was included in the experiment, this can still be indirectly assessed by scrutinising participants' produced V2 sentences for whether the word order following the verb adheres to the grammar of the language or not. If participants produced the constituents in the correct position, they have presumably recognised their relative positions in the tree. This should not be affected by the set types participants used to construct their sentences from. A mixed-effect logistic regression model was fitted to all produced V2 sentences that featured either a subject, direct object, indirect object or adjunct (simple or complex) in clause-initial position. The dependent variable indicated whether the postverbal word order was grammatical (=1) or not (=0). Sentences were considered grammatical when the constituents obeyed the following relative order of postverbal constituents: S > M > DO/IO > A. Sentences in which both a direct object and an indirect object followed the verb, were analysed as grammatical when the movement marker was right-adjacent to the verb. The model included SET TYPE (simple adjunct, complex adjunct, indirect object) as fixed effect, and by-participant and by-item random intercepts. By-participant random slopes were originally included in the model but subsequently dropped due to singular fit. SET TYPE was treatment coded with simple adjunct as baseline. The model indicated that participants produced the grammatical postverbal word order in sets with simple adjuncts significantly more frequently than expected by chance ($\beta = 1.36$, SE = .31, $p = 7.57\times^{-22}$). Participants did not use the grammatical postverbal word order significantly less often with sets containing either novel complex adjuncts ($\beta = -0.12$, SE = .30, $p = .69$) or novel indirect objects ($\beta = -0.21$, SE = .30, $p = .47$).

The main motivation for the present experiment was to improve participants learning of a V2 grammar compared to experiment 1. The hypothesis, however, made no reference to the results of experiment 1. Nevertheless, a comparison of experiment 1 and experiment 2 is necessary to evaluate the efficacy of the design changes. Figure 2.8 depicts the acceptance rate of different sentence types for experiment 1 and experiment 2 next to each other. Following the same analysis procedure as in the hypothesis-confirming analysis, I first compared how the ratings for V2-familiar sentences differ between experiment 1 and experiment 2. This is an important indicator for participants' learning success as those sentences correspond to the structures participants have been trained on. I fitted a mixed-effects logistic regression model to participants' judgements of V2-familiar sentences from experiment 1 and experiment 2. The model included EXPERIMENT as fixed effect and by-participant and by-item random intercepts. EXPERIMENT was treatment coded with experiment 1 as baseline. Investigating the simple effect of experiment, the model revealed that the acceptance rate of V2-familiar sentences in

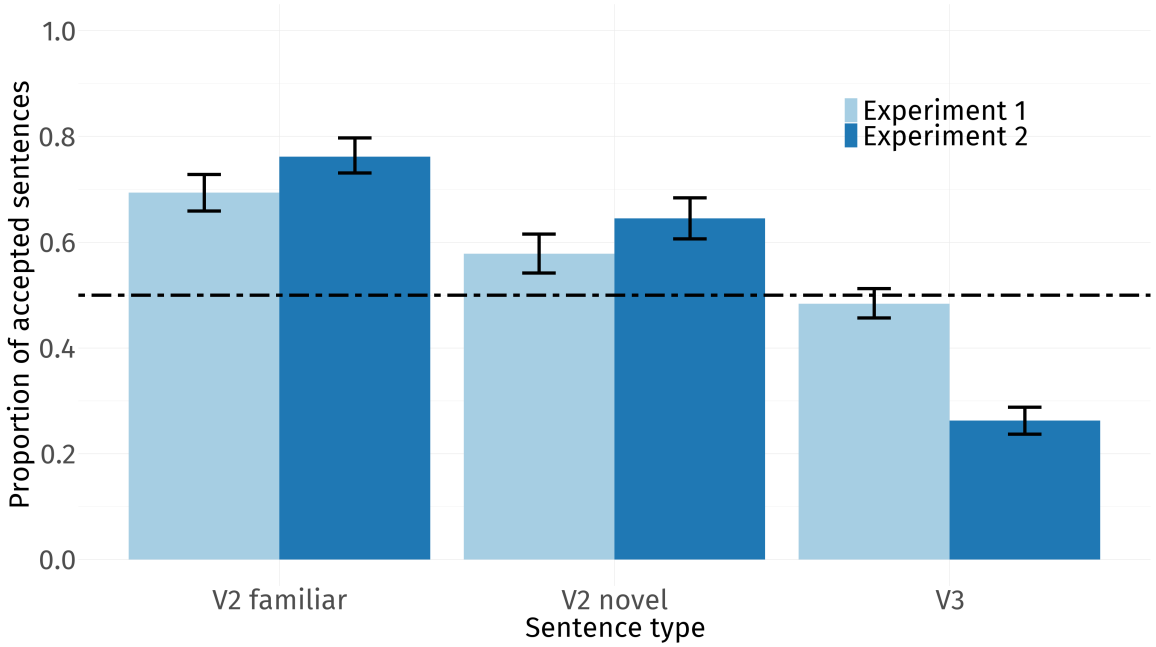


Figure 2.8: Acceptance rates of different sentence types in experiment 1 and experiment 2. Error bars indicate bootstrapped 95% confidence intervals. The dotted line represents the chance level. Participants in experiment 2 did not exhibit a significantly higher acceptance rate of V2-familiar and V2-novel sentences. However, the discrimination between V2-novel and V3 sentences — one of the central aspects of learning V2 as I defined it — was improved in experiment 2.

experiment 2 was not significantly higher than that in experiment 1 ($\beta = .47$, SE $= .29$, $p = 0.11$).

In a next step, I examined how experiment 1 and experiment 2 differ for V2-novel and V3 sentences. I fitted a mixed-effects logistic regression model to the V2-novel and V3 data. The model included SENTENCE TYPE and EXPERIMENT as fixed effects as well as an interaction term of both fixed effects. The model further included by-participant and by-item random intercepts as well as by-participant random slopes for SENTENCE TYPE. Both fixed effects were treatment coded with V2-novel and experiment 1 as reference level, respectively, as reference levels. First, I examined the simple effect of EXPERIMENT to asses whether the acceptance rates for V2-novel sentences differ. The model did in fact not show a significant difference between the ratings for V2-novel sentences in both experiments ($\beta = .38$, SE $= .27$, $p = .15$).

I also studied how the discrimination between V2-novel and V3 sentences differed between experiments by looking at the interaction of SENTENCE TYPE and EXPERIMENT. As can be seen in Figure 2.8, the ratings for V3 sentences are lower in experiment

2, suggesting a better discrimination. This was confirmed by a significant interaction between V3 and experiment 2 ($\beta = -1.68$, SE $= .34$, $p = 1.04 \times 10^{-6}$).

2.4.4 Discussion

Experiment 2 investigated the same hypothesis as experiment 1; that is, V2 can be learnt in an ALL experiment. Although the materials remained unaltered from experiment 1, two crucial changes were made in comparison to experiment 1. First, production trials were included in the training, following findings of Hopman & MacDonald (2018). As a result, a production test was also included in the testing phase of the experiment. The second change was the tested population. Participants were recruited through Prolific instead of MTurk. This change was motivated by recent findings indicating a generally better performance of Prolific users in behavioural experiments (Peer et al. 2021, Uittenhove, Jeanneret & Vergauwe 2023). The eventual goal of the altered experiment design was to boost the performance of participants in comparison with experiment 1.

The results provide strong evidence for the investigated hypothesis. Participants' judgements exhibit the same pattern as in experiment 1 in that all of my predictions were confirmed: The high acceptance rate of V2-familiar sentences indicates that participants have learnt the structures in the input. Furthermore, the high acceptance rate of V2-novel sentences suggests that participants had extrapolated the V2 rule to novel sentence types. This is further supported by the low ratings for V3 sentences showing that participants were able to discriminate between novel grammatical and ungrammatical structures.

When it comes to the production data, the results did not fully match my predictions. However, as I will argue, they still provide support in favour of the hypothesis. The frequency of V2 sentences varies for different sentence types, contrary to my prediction. The analysis has shown that the frequency with which verbs are placed in second position in productions is significantly above chance. Even though sentences with complex adjuncts and indirect objects had significantly fewer verbs in second position than sentences with simple adjuncts, the overall proportion is still very high (cf. Figure 2.5). That is, participants still extrapolated the position of the verb to sentences with novel constituents. In fact, this pattern matches the judgement data perfectly. As shown in the exploratory analysis, the acceptance rate of V2-familiar sentences is significantly higher than the one of V2-novel sentences. Further support that V2 was learnt comes from the analysis of postverbal word order. Participants appear to have assigned the hierarchical structure commonly associated with V2 languages to the input (i.e. sentence-medial adjuncts adjoined to VP/*v*P). This could indicate that verbs undergo movement to the left periphery — a defining feature of V2.

The second prediction for the production data was also not confirmed. I predicted that participants should produce V2 sentences with novel constituent types in clause-initial position (as opposed to V2 sentences with familiar clause-initial elements) at higher than chance level. The statistical analysis revealed though that participants produced V2 sentences with clause-initial novel constituents below chance level. Recall that the

production test preceded the judgement task and exclusively previously unseen lexical items were used. In this light, the fact that participants placed novel constituents in clause-initial position at all is worth underscoring. In fact, the prediction might have been too strong. There is no evident reason as to why participants should use a novel constituent type more frequently than others. Studies have shown that participants probability match their input under certain circumstances (Hudson Kam & Newport 2005, 2009, Smith & Wonnacott 2010). That is, participants reproduce the input frequency of constructions more or less veridically in experiments. If participants were probability matching, one would expect only very few initial novel constituent types. Consequently, the fact that novel constituent types account for over 10% of all constituents placed in clause-initial position in V2 sentences supports the overall observed extrapolation. This conclusion would still hold, if processes other than probability matching were at play here. For instance, it is conceivable that participants have formed ad hoc strategies during training to account for different structures. These strategies would not licence structures with clause-initial novel constituent types. Hence, the production of sentences like these provides strong evidence that participants extrapolated the flexibility of clause-initial position to novel structures.[32]

Taken together, the results of experiment 2 confirmed the findings of experiment 1 as to the learnability of a V2 language in ALL experiments. However, the main motivation for altering the initial experiment design was to increase the learning performance of participants. The exploratory analysis has shown that the acceptability rates of V2-familiar and V2-novel sentences in experiment 2 are, albeit numerically higher, not statistically different from those in experiment 1. At the same time, participants in the second experiment discriminated significantly better between V2-novel sentences and ungrammatical V3 sentences. This contrast constitutes an important finding because it demonstrates that the changes to the experiment design were indeed effective. Although the changes did not improve the degree to which participants extrapolated V2 to novel structures, they still improved learning. In other words, learners gained confidence in the position of the verb. This finding poses the question whether training with production trials, the different population or a combination of both contributed to the improved performance. Unfortunately, the current design does not allow me to identify the extent to which the two factors contribute to the improved learning performance. A comparison with Rebuschat (2008) suggests, however, that the production component affected learning positively at least. The acceptance rate for V3 sentences in the current experiment (26.2%) approaches the one reported by Rebuschat (2008) closely (22.9%, cf. Table 2.2). Seeing that both experiments involved some kind of production during training, the parallel between the two experiments could be explained by the production component of both. This would also shed new light on the relatively high acceptance

[32]The distribution of clause-initial constituents is inconclusive as to whether participants actually probability matched. While subjects occur with approximately the same frequency as in the input in initial position (36.0%), direct objects were clearly overproduced (49.8%). Participants might have extended (i.e. generalised) the contexts where direct objects are used — a process that has been observed in previous work (e.g. Hudson Kam & Newport 2009, Ferdinand, Kirby & Smith 2019, Keogh, Kirby & Culbertson 2022).

rate for V3 sentences observed in experiment 1. In the corresponding discussion, two potential causes for the high ratings were identified. On the one hand, participants underwent less training in my experiment compared to Rebuschat (2008). On the other hand, the language in my experiment exhibited more differences from participants' native language in terms of permissible clause-initial elements. In the light of the findings of experiment 2, the lack of a production during training could be a third explanation for the low ratings.

2.5 General discussion

The primary objective of this thesis is to investigate how variability in the input affects the acquisition of a V2 language. In Chapter 1, artificial language learning (ALL) has been identified as a suitable experimental methodology for addressing this research question. The goal of the current chapter was therefore to establish an experimental design that enables participants to learn a V2 language. From the discussion of previous work that has employed either semi-artificial of fully-artificial V2 languages (without investigating V2 itself), valuable insights could be gained. First and foremost, V2 can indeed be learnt in an ALL experiment. However, I also identified four design aspects requiring revision. First, participants should be taught the language incidentally to ensure that participants use solely their intuitions at test. Second, the training items should all be semantically plausible as implausible items might slow down learning. Third, the syntax should be kept minimalist. That is, just a single verb position should be included unless the research question explicitly dictates otherwise. Finally, the way in which learning a V2 language is measured should be revised. It was argued that participants' generalisation of the flexibility of the clause-initial position to novel constituent types should be assessed. The efficacy of the suggested modifications were then tested in an ALL experiment. Participants were taught a semi-artificial language consisting of English vocabulary and a V2 syntax and subsequently tested on their knowledge of the language. All predictions were confirmed. Participants learnt V2 structures with familiar constituent types in clause-initial position. Furthermore, participants generalised XP-fronting to the clause-initial position to previously unseen constituent types, while rejecting ungrammatical V3 sentences. These results thus add to earlier findings showing that V2 can be learnt in an ALL experiment. A descriptive comparison with the results of Rebuschat (2008) revealed a similar acceptance rate of V2 sentences with familiar constituents in clause-initial position. However, the ratings for V3 sentences were significantly lower in Rebuschat (2008). Although this might be interpreted as ineffectiveness of the modifications to the design, I argued that the opposite is the case. The language learnt by participants in my experiment deviated more from participants' native language due to the exposure to uncommon object-initial sentences. The fact that participants achieved a similar degree of learning for V2 sentences with familiar clause-initial constituents (in face of the increased language difficulty) points to the effectiveness of the introduced changes.

Although experiment 1 successfully demonstrated the usefulness of the design changes, two further changes were suggested to improve the learning performance. The first concerned the inclusion of a production component during training, whereas the second

pertained to the tested population. Instead of recruiting participants from MTurk, Prolific should be used for recruitment. The first experiment was thus repeated with the suggested changes. Akin to experiment 1, the judgement data showed that participants learnt the input structure and extrapolated XP-fronting to novel constituent types. At the same time, ungrammatical sentences exhibited a lower acceptance rate than V2 sentences with familiar and novel constituent types in clause-initial position. Although not all predictions were confirmed for the production data, there was still strong evidence that participants learnt the language. The comparison of the judgement data from experiment 1 and experiment 2 revealed that the only statistically reliable difference between the two experiments lies in the ratings for V3 sentences: The acceptance rate was significantly lower in experiment 2. That is, the inclusion of production trials and a changed population increased participants confidence in the position of the verb.

The two experiments combined demonstrate that ALL constitutes a suitable methodology for investigating V2. Participants are able to learn the language to a degree that allows studying the effect of variability in the input on the acquisition of V2. This is reflected in participants' ability to discriminate between grammatical and ungrammatical patterns. My findings add to earlier studies that have used a V2 language for investigating different questions. The two experiments, however, provide a stronger test of learning V2. Previous studies only tested participants knowledge with structures that were familiar to participants. That is, by showing that participants extrapolate XP-movement to novel types of constituents, compelling evidence is provided that participants learn a genuine V2 grammar and not a grammar that merely licences subject-, object- and adjunct-initial structures.

The results of the two experiments have important implications for experimental designs for studying V2. The first concerns the type of the language. Even though artificial and semi-artificial languages can be learnt in an experimental setting, semi-artificial languages offer certain advantages over fully-artificial languages. On the one hand, participants can focus exclusively on learning the syntactic patterns of the language. On the other hand, a broader range of lexical items can be used. By increasing the number of lexical items (compared to a fully-artificial language), chances decrease that participants derive incorrect lexically conditioned rules. Although one might object to the use of semi-artificial languages on grounds of their close resemblance to participants' native language, the results of the two experiments contradict such a conclusion. The used V3 structures are very similar to topicalisations in English. If participants used their native language as basis for their judgements, V3 sentences should have received significantly higher acceptance rates than they actually did. There is also independent evidence suggesting that such a stance is unsubstantiated. Culbertson & Adger (2014) taught English-speaking participants an English-based language in which modifiers follow the noun. Participants were exposed to only one modifier at a time during training. At test, participants had to infer the order of two or three modifiers. Interestingly, participants did not simply reproduce the English word order (i.e. noun-adjective-numeral-demonstrative) but rather homomorphic orders (i.e. demonstrative-numeral-adjective-noun). When later repeated in a fully-artificial language learning

experiment, the same results were obtained (Martin et al. 2020). In other words, the semi-artificial language did not confound the results. Hence, semi-artificial language can be used for studying V2.[33]

The second implication also relates to the nature of the language. In the discussion of the results, it has become evident that multiple verbal positions slow down the learning process, especially when participants are trained on a semi-artificial language. The language should thus be kept as simple as possible. This also includes the use of exclusively semantically plausible sentences.[34]

Another implication pertains to the manner in which participants are taught the language. The contrast between experiment 1 and experiment 2 indicates that the introduction of a production component fosters participants' learning. The benefits of a production component lie not so much in a better generalisation of XP-movement but rather in a better discrimination between ungrammatical sentences and structures with unfamiliar clause-initial constituents.

Finally, the experiments also highlighted that the participant pool has to be chosen very carefully. If experiments are not run in-person in a lab but online, not all platforms might be equally suitable. Following previous research suggesting that MTurk users are less suitable for experimental work, the participant pool was changed from MTurk users to Prolific users. When all these insights are implemented, V2 can be studied experimentally.

2.6 Summary

The purpose of this chapter was to develop an experimental design that enables the investigation of V2 with artificial language learning (ALL) experiments. After reviewing previous work that used artificial V2 languages in their experiments, potential modifications of previous designs were identified. Albeit showing that V2 can indeed be learnt in an experimental setting, previous experimental designs included features that could potentially hinder learning. The efficacy of the alterations was tested in two experiments using a semi-artificial language. The results of the experiments show that participants' learning improved, even though this was somewhat obscured by the use of a language with a broader variety of clause-initial constituent types. Additionally, the results underscored the benefits of including production trials in the training as well as the need to choose the tested population carefully. With the experimental design in place, the research question can now be investigated. As the results of the next chapters will show, the hypothesis developed in Chapter 1 needs to modified to incorporate the role of variability in grammatical categories.

[33] Note that the experimental results outlined in the subsequent two Chapters indicate that participants' L1 does in fact confound the results, at least to some extent. As already mentioned above, this finding will prompt a revision of the initial hypothesis formulated in Chapter 1 such that variability in grammatical categories will be taken into consideration for the learnability of a V2 grammar.

[34] It might, however, be desirable to examine the contribution of alternating verb positions as found in SOV languages with V2 orders in future work.

CHAPTER 3

COMPARING THE EFFECTS OF VARIABILITY IN THE CLAUSE-INITIAL POSITION

3.1 Introduction

The main goal of this thesis is to investigate the relation between the loss of V2 and the variability in the clause-initial position. According to the hypothesis developed in Chapter 1, lower variability in the initial position should lead to the loss of V2: Without such variability, learners will fail to generalise that no fixed association exits between the clause-initial position and a particular grammatical function (e.g. subjects or objects). To investigate this hypothesis, an experimental design was devised in Chapter 2 that enables the study of V2 with artificial language learning (ALL) experiments. The present chapter reports the results of an ALL experiment (i.e. experiment 3) scrutinising the hypothesis formulated in Chapter 1 by using the experimental design developed in Chapter 2. Three conditions were compared which differed only with respect to the proportion of different clause-initial constituent types. Participants in the first condition were trained on a semi-artificial language that exhibits the highest variability in the clause-initial position. That is, subjects, direct objects and adjuncts were realised equally frequent in the initial position. The two remaining conditions were characterised by less variability: Both conditions are similarly skewed, but the dominating elements differ; either adjuncts or objects account for the most frequent clause-initial element. Recall from Chapter 2 that the participants in both experiment 1 and experiment 2 were trained on a language that featured all constituent types with the same frequency in clause-initial position. In other words, the experiments already tested one of the conditions. For this reason, the results of experiment 2 (cf. §2.4) — whose design was identical to experiment 3 — were compared with the two skewed conditions. Experiment 1 was unsuitable due to the differences in the design. That is, only two additional conditions were run.

The chapter further presents the results of a large-scale corpus study that investigated the distribution of clause-initial constituents in German. The study thereby deviates from most previous work, which examined only small data sets. Even though the focus of this thesis lies on experimental work, corpus work can provide further valuable evidence

for the question at hand. If present-day V2 languages are characterised by a highly variable clause-initial position, such distributions could serve as additional evidence in favour of the hypothesis investigated here. As this chapter will show, however, the experimental results do merely provide partial support for my hypothesis. Although participants exposed to a high variability in clause-initial grammatical functions showed better learning than the learners in the condition where objects dominated, learners in the adjunct-dominant condition outperformed both conditions. Two (mutually non-exclusive) interpretations will be offered: First, the apparent learning advantage of the adjunct-dominant condition could be attributed to a facilitating influence from participants' L1. Second, a large amount of adjuncts could be beneficial for learning a V2 grammar. This can be explained if learners are not sensitive to variability in clause-initial grammatical functions but rather grammatical categories — hence requiring a revision of the hypothesis developed in Chapter 1. Crucially, a redefinition of the relevant domain of variability does not constitute an issue: As stated in §1.5.2, the focus on grammatical functions was merely a stipulation. The corpus corpus findings also lend support to the second interpretation in that a high proportion of clause-initial adjuncts was found for German.

The main body of the chapter (i.e. the results of the experiment and the corpus study) has been accepted for publication in *Journal of Historical Syntax*. The paper has been previously published as the following co-authored preprint and has benefited from the comments of two anonymous reviewers:

Marc Meisezahl, Simon Kirby & Jennifer Culbertson. 2023. Variability and learning in language change: The case of V2. OSF Preprints. doi:10.31219/osf.io/c6gbp

CRediT author statement Marc Meisezahl: Conceptualisation, Methodology, Software, Formal analysis, Investigation, Data Curation, Writing – Original Draft, Writing – Review & Editing, Funding acquisition. Simon Kirby: Conceptualisation, Methodology, Writing – Review & Editing, Funding acquisition. Jennifer Culbertson: Conceptualisation, Methodology, Writing – Review & Editing, Funding acquisition.

3.2 Variability and learning in language change: The case of V2

ABSTRACT The loss of V2 has received considerable attention in the past with some theories linking it to learning (e.g. Lightfoot 1999, Yang 2002). Here, we use artificial language learning experiments to test, in a controlled setting, what factors affect learning of V2. Specifically, we build on previous work demonstrating a general beneficial effect of input variability. We explore the role of variation in clause-initial constituents by comparing artificial languages that differ both in the kinds of grammatical functions that tend to appear in initial position, and the level of variability present. We find that these different distributions of clause-initial constituents indeed affect V2 learning outcomes. However, contrary to our predictions, a language with the highest level of variability is not the best learnt. Rather, a language containing many adjunct-initial sentences was learnt best. We discuss the possibility that a high quantity of clause-initial adjuncts is in fact important to acquiring V2 grammars in natural language. We find further support for this in corpus data indicating a high proportion of adjunct-initial sentences in stable V2 languages and a low proportion in languages that had been in the process of losing V2. We also discuss the role of variability in grammatical *categories* rather than roles, which might give languages with many clause-initial adjuncts an advantage. Taken together, our findings establish the first evidence for a causal link between the reduction of evidence and the loss of V2.*

1 Introduction

Most modern Germanic languages are verb second (V2) languages. That is, the verb is obligatorily realised in the second position of a clause and no restrictions apply as to the grammatical function or category of the clause-initial constituent (Holmberg 2015). For instance in German, subjects (1a), objects (1b), adjuncts (1c) and past participles can all occupy the initial position (1d), but verbs must be in second position.

(1) a. Die Maus **geniesst** eine Scheibe Raclettekäse auf der Alm.
the mouse enjoys a slice raclette cheese on the alp
'The mouse enjoys a slice of raclette cheese on the alp.'

*We are indebted to Mora Maldonado for helping with the experiment design, the analysis and providing parts of the code. Furthermore, we thank Alexander Martin and the audiences of CONSOLE 30 in Nantes and DiGS 23 in New York and two anonymous reviewers for valuable input. This project received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 757643), and from ESRC grant ES/R011869/1.

b. Eine Scheibe Raclettekäse **geniesst** die Maus auf der Alm.
a slice raclette cheese enjoys the mouse on the alp
c. Auf der Alm **geniesst** die Maus eine Scheibe Raclettekäse.
on the alp enjoys the mouse a slice raclette cheese
d. Genossen **hat** die Maus eine Scheibe Raclettekäse auf der Alm.
enjoyed has the mouse a slice raclette cheese on the alp

While on the surface, many English sentences also follow a V2 pattern, subjects must occupy the preverbal position (2a), and additional elements in the left-periphery lead to V>2 orders (2b).

(2) a. *A slice of raclette cheese **enjoys** the mouse on the alp.
b. On the meadow, the mouse **enjoys** a slice of raclette cheese.

This reflects change: Earlier stages of English followed a word order pattern akin to (1) (van Kemenade 1987, Fischer et al. 2001), a pattern of change that is attested in a number of other languages (Willis 1998, Meelen 2016, Wolfe 2018). The loss of V2 has been the subject of longstanding research, with some theories tying it to learning. Like any other feature of language, evidence for V2 has to be sufficient in the input learners receive, otherwise it will not be acquired (Lightfoot 1999, Yang 2000). For V2, it has been claimed that exposure to sentences with verbs in second position does not on its own suffice. Rather, non-subject-initial sentences, like the ones in (1), form a crucial part of the evidence for V2 (Yang 2000). This is supported by historical data showing a link between the increase in subject-initial sentences and loss of V2 (e.g. Roberts 1993: 199). Intuitively, if a large proportion of the sentences in the input involve SV orders, then the learner may acquire a grammar similar to modern English, rather than V2. In principle, exposure to a preponderance of sentences with *any* particular constituent type in first position (e.g. subject, object, etc.) could lead the learner to an analysis that favours a non-V2 grammar.

In this paper, we use a novel experimental method to explore the role of variation in the initial constituent in learning V2. Specifically, we ask whether learning V2 in a miniature artificial language is affected by the level of variability in the grammatical roles (i.e., subject, object and adjunct) of clause-initial constituents. We are specifically interested in the way participants generalise from the input. Forming abstract grammatical representations from the input constitutes a central task in language learning. In the case of V2, learners need to build representations without a fixed mapping of grammatical roles to the clause-initial position. If more variability increases the likelihood of generalisation — i.e., to novel types of constituents — and less variability decreases this likelihood, we will have the first direct evidence for a causal link between the frequency of non-subject-initial sentences and the loss of V2. If a learner fails to generalise to novel initial constituents but instead learns a more constrained grammar, where a specific grammatical role or set of roles appear in initial position, then the first steps toward the loss of V2 have essentially been taken. When considered on a larger timescale, even a weak tendency toward learning a more constrained grammar can

be amplified through cultural transmission (Kirby, Cornish & Smith 2008, Kirby et al. 2015) with the result that V2 is lost at the population-level.

The results of our experiment suggest that the learning and hence generalisation of a V2 grammar is indeed affected by variation, however we found that learning is best when the language exhibits a large amount of *adjuncts* in clause-initial position and worst when objects dominate the clause-initial position. Although this does not exactly fit our initial hypothesis, which concerned variation in grammatical functions, our results still provide evidence for a fostering effect of variability on learning. Specifically, this may indicate that instead of a high variability of grammatical functions, a high variability of grammatical categories (e.g. DPs, PPs & AdvPs) in the clause-initial position is what fosters generalisation. This finding is further supported by a large-scale corpus study on the distribution of clause-initial constituents in German. In line with previous studies on Germanic and Romance, we find that adjuncts account for the most frequent constituent type in clause-initial position after subjects.

This paper is structured as follows: We first motivate our hypothesis (§2). We will then report the results of our artificial language learning experiment (§3). In §4, we will present a large-scale, multi-corpus analysis of variation in the clause-initial constituent in a natural V2 language, German. We will conclude with a discussion of what our findings mean for research on the historical loss of V2 in §5.

2 The role of learning in the loss of V2

V2 is a cross-linguistically rare phenomenon (Holmberg 2015: 343).[1] There are some languages (including most Germanic languages other than English) which have exhibited relatively strict V2 since their earliest records, in some cases developing even stricter V2 order over time (Eythórsson 1995, Axel 2007, 2009, Axel-Tober 2018, Þorgeirsson 2012). There are other languages which have had and lost, to varying degrees, their V2 status (including English van Kemenade 1987, Roberts 1996, the Romance languages Benincà 1995, Wolfe 2018 and Welsh Willis 1998, Meelen 2016). Why is V2 rare, and what leads some languages to retain, and even strengthen V2, while others lose it?

One possibility is that the rarity and fragility of V2 is due to the kind of evidence that is needed for learners to acquire it. In a prominent account, Yang (2000, 2002) argues that the loss of V2 is tied to changes in the linguistic input of learners. For example, looking at the case of French, Yang (2000) argues that to retain a V2 grammar, unambiguous evidence to support V2 (i.e. OVS and XVSO sentences, must outnumber the evidence against V2 (i.e., SXVO and XSVO). Crucially, because Middle French was a *pro*-drop language, non-subject-initial V2 sentences could not provide unambiguous evidence for V2: a V2 analysis i.e., [X V *pro*], or an SVO analysis i.e., [X *pro* V], are both possible. Yang (2000) uses counts from Roberts (1993: 148, 155) to argue that in Middle French, unambiguous V2 sentences decreased to such an extent that a SVO grammar gained an advantage over a V2 grammar. The cue-based learning model of Lightfoot (1999, 2006) also takes a learner-centred approach to the loss of V2. In this model, cues

[1] To the best of our knowledge, 21 languages have been categorised as V2.

— $_{CP}$[XP $_{C}$V…] in the case of V2 (Lightfoot 2006: 86) — need to be sufficiently expressed in the input to ensure successful learning (Lightfoot 2006: 82). Both of these accounts suggest that sufficient unambiguous evidence for V2 must be present for successful acquisition. But they also highlight the role of variability in sentence structures as a key aspect of evidence for V2. For example, it is not just the absence of V3 sentences that matters, but the presence of OVS and XVS alongside SVO sentences.

There is independent evidence from various domains suggesting that greater variability can in some cases benefit learning (see Raviv, Lupyan & Green (2022) for a recent review). For example, Gómez (2002) and Gómez & Maye (2005) show that learning of non-adjacent dependencies is successful only when there is sufficient variation in the elements that occur between two dependents. For example, learners exposed to sequences of the type 'aXc' only learn that 'a' elements must be followed by 'c' elements when there are sufficiently many different 'X' elements. Here, variability helps learners to focus on key patterns of interest, and rule out irrelevant information: variability in the intervening elements helps learners to move from a focus on transitional probabilities between adjacent elements (e.g. 'aX' or 'Xc') to the non-adjacent dependency. The elements in these experiments could in principle represent grammatical categories (e.g. pronoun + verb + agreement marker), constructions (e.g. *be* + verb + *-ing*), or grammatical functions (Gómez 2002: 431). While the current study does not focus on dependencies, the intuition remains the same: variability helps draw the learner's attention to critical constraints — here the position of verb — and provides evidence for what is unconstrained — the sentence-initial position. In line with Yang (2000) and Lightfoot (1999, 2006), we specifically pursue the idea that more variation *in the clause-initial constituent* provides more robust evidence for generalised XP-movement and a V2 grammar. In particular, we predict that V2 grammars are learnt best when variation in the types of grammatical roles (i.e., subjects, objects and adjuncts) of clause-initial constituents is highest. The degree to which participants generalise XP-movement will affect the preservation of V2 in the grammar. A weaker generalisation will eventually lead to the loss of V2 as constraints on the clause-initial position are amplified through cultural transmission.[2,3] We test this prediction using an artificial language learning experiment.

[2]See also Cournane & Klævik-Pettersen (2023) for another account of the loss (and rise) of V2 that focuses on the role of learning. Their account highlights the conservative nature of learners when it comes to the acquisition of syntactic structures. Note that the account of Cournane & Klævik-Pettersen (2023) and our account are not necessarily in conflict but can complement each other.

[3]It should also be noted that the positive effects of variation in this context do not contradict the potential negative effects of variation elsewhere in learning. For example, when a fixed rule must be learned, then variation in the form of exceptions can be problematic for learning. For example, according to the Tolerance Principle (Yang 2016), rule learning is sensitive to a specific threshold of exceptions. Specifically, the Tolerance Principle predicts that if the number of exceptional forms (e.g. irregular past tense forms) remains below a threshold, a rule (e.g. past tense) will be considered productive by learners. Crucially in the context of V2, the rule learners need to acquire is that the mapping between first position and grammatical role is *not* fixed. Thus in this case, more variability provides more evidence.

3 Artificial language learning experiment

To test our hypothesis experimentally, we conducted an artificial language learning (ALL) experiment. ALL studies allow researchers to create miniature linguistic systems in which variables of interest can easily be manipulated and variables not of interest controlled (for a review see Culbertson & Schuler 2019). It has been shown that artificial languages are learnt in similar ways to natural languages (Ettlinger et al. 2016), and there is substantial evidence that learners' preferences in ALL studies align with linguistic typology (Culbertson 2012, 2017, 2023, Culbertson & Schuler 2019). Consequently, participants' behaviour in experiments provides an important source of evidence from which to draw conclusions about the link between learning and natural language structure. Previous studies have suggested that it is possible to learn V2 in an artificial language (Getz 2018, 2019, Rebuschat & Williams 2012, Tagarelli et al. 2016, Ruiz, Tagarelli & Rebuschat 2018). In our study, participants learn a novel miniature artificial language involving English lexical items which conform to a (non-English-like) V2 grammar. The verb always comes second, but our study design manipulates the distribution of clause-initial elements participants are exposed to. We included three conditions: a uniform condition in which subjects, direct objects and adjuncts occurred equally frequently in clause-initial position and two conditions with skewed distributions where either adjuncts or objects accounted for the majority of clause-initial constituents. We predict learners in the uniform condition will be more likely to acquire a V2 grammar compared to the skewed conditions. Importantly, we measure learning V2 in terms of the critical feature of V2 languages: generalisability of clause-initial position (equivalent to XP-fronting).

3.1 Methods

The design, the hypotheses, predictions and analyses were preregistered prior to data collection. We implemented the experiment using the JavaScript library jsPsych (de Leeuw, Gilbert & Luchterhandt 2023). All materials are available online.

3.1.1 Participants

314 participants were recruited online using Prolific. By using in-built Prolific filters, the participant pool was restricted to United States nationals, who are monolingual speakers of English that were also raised monolingually.[4] We also used Prolific filters to exclude participants whose subject at university was English literature, English language or languages more broadly. Finally, only participants with an approval rating of 95% or higher were invited to participate. Following our pre-registered exclusion criterion, 82 participants were excluded due to low performance in the first half of the experiment (cf. §3.1.3) and two participants had to be excluded due to knowledge of a V2 language (as determined by a post-experiment questionnaire). Data analysis therefore included

[4]The experiment received ethical approval from the ethics board of the Linguistics and English Language department at the University of Edinburgh (180-2021/2). All participants gave informed consent before their participation.

74 participants in the uniform condition and 78 participants in each of the two skewed conditions.

3.1.2 Materials

Stimuli sentences were constructed using a semi-artificial language. The vocabulary of the language consisted of English lexical items, but the word order followed a non-English-like V2 pattern. All sentences were comprised of a verb, a subject, an object, an adjunct, and an additional adverb. The verb was always in second position, with either a subject, an object, or an adjunct in initial position. The adverb served as an additional cue to the non-English-like structure of the language, cf. (3). It always appeared to the right of the finite verb (unlike in English), indicating movement of the verb out of the VP/*v*P to a higher functional projection (Vikner 1995, Waldmann 2008, Westergaard 2009a). Using English lexicon items allowed us to train participants on the grammar in a short time (given the length and complexity of the sentences), and allowed us to control for lexical novelty (as described below) in testing whether participants generalise XP-fronting to novel types of constituents in each condition. A number of other studies have demonstrated that results obtained with semi-artificial languages can be replicated with a fully-artificial language (Culbertson & Adger 2014, Martin et al. 2019, 2020).

The distribution of elements (subject, object, adjunct) in first position depended on the condition. In the *uniform* condition, all three were equally likely to appear in initial position. In the *object-dominant* condition, objects were more likely to appear in initial position. In the *adjunct-dominant* condition, adjuncts were more likely to appear in initial position. Because we use an English lexicon, and English is SVO, we did not run a subject-dominant condition. Participants were randomly assigned to a condition.

We created 30 unordered sets of constituents used for training. For each set, three different sentences — subject-initial, object-initial and adjunct-initial — were created yielding 90 sentences in total. This process is illustrated in (3) (bold, italics and underline used here for illustrative purposes only). Each constituent was a phrase made up of a single word or two words. Constituents in each set were unique, and no constituent was used in more than one set.

(3) {revises, in Boston, a novel, the author}

a. *The author* **revises** eventually <u>a novel</u> in Boston.
b. <u>A novel</u> **revises** *the author* eventually in Boston.
c. In Boston **revises** *the author* eventually <u>a novel</u>.

Subjects were always animate and objects inanimate. Subjects were DPs or proper nouns, objects were DPs. To further facilitate identification of grammatical roles, only verbs denoting irreversible actions were included. Adjuncts were temporal or locative adverbs, PPs or few adverbially used DPs.

For the uniform condition, we used all subject-initial, object-initial and adjunct-initial sentences created from the 30 unordered sets. That is, each constituent type occurred equally frequent in clause-initial position in the training. By contrast, for the skewed

Condition	Subject-initial	Object-initial	Adjunct-initial
Uniform	30	30	30
Object-dominant	18	54	18
Adjunct-dominant	18	18	54

Table 3.1: Distribution of subject-initial, object-initial and adjunct-initial sentences in each of the three conditions.

conditions, the dominant element accounted for 60% of all sentences, whereas non-dominant elements accounted for 20% each (Table 3.1). We chose 60% as frequency for the dominant constituent type as this lies within the range of the dominant constituent in V2 languages (cf. §4). As a result of the skew, only a subset of all possible combinations could be included and some sentences were repeated. For eight of our thirty unordered sets, we replaced one of the sentence variants with a non-dominant constituent in clause-initial position with the variant featuring the dominant constituent (i.e., either adjunct or object) in clause-initial position. For example, participants in the adjunct-dominant condition could have seen (3a) once and (3c) twice during training. We repeated sentences rather than including new ones in order to control lexical variation across conditions. Sets were randomly selected for repetition for each participant. We assigned each sentence of the same set to three different blocks. The distribution of clause-initial elements in each block adhered to the same overall pattern (i.e. uniform or skewed). For each condition, three training lists were created from the three different blocks by Latin Square. The order in each block was randomised for each participant.[5] The first training block will be used for the reading trials and the second and third block for the production trials (cf. §3.1.3).

For the testing phase, two different sets of materials were constructed. The first type was used for production testing, and consisted of unordered sets of constituents that were presented to participants as buttons they could choose to create a sentence (see §3.1.3). Each set could contain the same types of constituents featured in training (i.e., subject, a direct object, adjunct, adverb) or they could contain one of two novel constituent types. In *complex adjuncts* trials, adjuncts comprised of three words (i.e., containing an additional determiner or modifier as in (4b)) rather the simple adjuncts used in training (as in (4a)). These trials allow us to test whether participants are sensitive to constituent length when they make generalisations about what can be fronted. In the *indirect object* trials, an indirect object replaced the adjunct (4c). These trials allow us to test whether participants generalise XP-fronting to novel constituent types.

(4) a. {the driver, delivers, grumpily, the food, this afternoon}
b. {Jayden, sweeps, halfheartedly, the floor, in the bathroom}
c. {Charles, suggests, cheekily, a whiskey, to the friend}

[5]Examples of a training set for each condition can be found here.

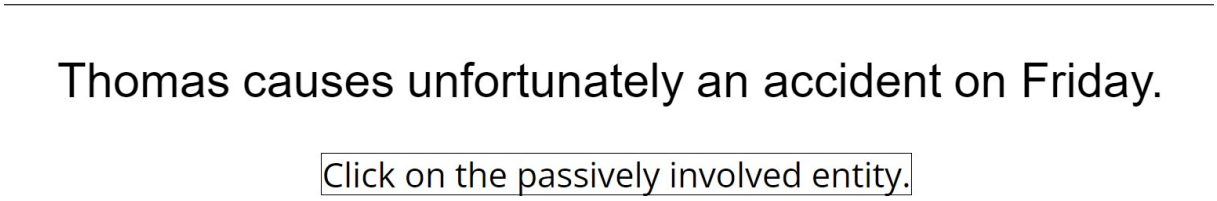


Figure 3.1: Example trial for reading task during training. Constituents were revealed one at a time, i.e., *Thomas*, *causes*, *unfortunately*, *an accident* and *on Friday*. Participants were then prompted to identify a particular constituent by clicking on it.

Four sets of each trial type — i.e., seen in training (i.e., with a simple adjunct), complex adjunct, or indirect object — were constructed, for twelve total sets. All constituents apart from adverbs were novel in the sense that they were not words or phrases seen in the training stimuli. Adverbs could be repeated from the training stimuli.

The second set of testing items was constructed for use in the sentence judgement phase (see §3.1.3). These items were created by crossing two factors: VERB POSITION (V2 or V3) and INITIAL CONSTITUENT (simple adjunct, complex adjunct, direct object, or indirect object). (5a) and (5b) exemplify the V2 sentences with initial indirect objects and complex adjuncts, respectively. Similarly, V3 sentences with initial indirect object and complex adjunct are illustrated in (6a) and (6b). For each factor combination, four sentences were created. We applied the same construction criteria on the different constituents as described above. Again, only adverbs were repeated, all other constituents involved new lexical items.

(5) a. To the congregation **shows** *the priest* silently the candle.
b. In late April **regrets** *the politician* openly his misconduct.

(6) a. To the doctor *the patient* **describes** precisely the pain.
b. At the moment *the referee* **verifies** briefly the decision.

3.1.3 Procedure

Participants accessed the experiment through a web browser on their personal computer or laptop. At the start of the experiment, participants were informed that they would be learning a recently discovered dialect of English that differs from other varieties of English in greater flexibility of the word order. The experiment was divided into a training phase and a testing phase. Participants were assigned randomly to one of three training lists per condition. The training phase was comprised of two parts: sentence reading and production. Before the reading trials, participants were told they would see sentences and be asked to identify either the actor of the action (i.e., subject), the passively involved entity (i.e., object), the action (i.e., verb) or the time/location of the action (i.e., adjunct). Then, on each trial, a sentence was revealed, one constituent at a time, with a delay of 500ms between constituents in order to give participants

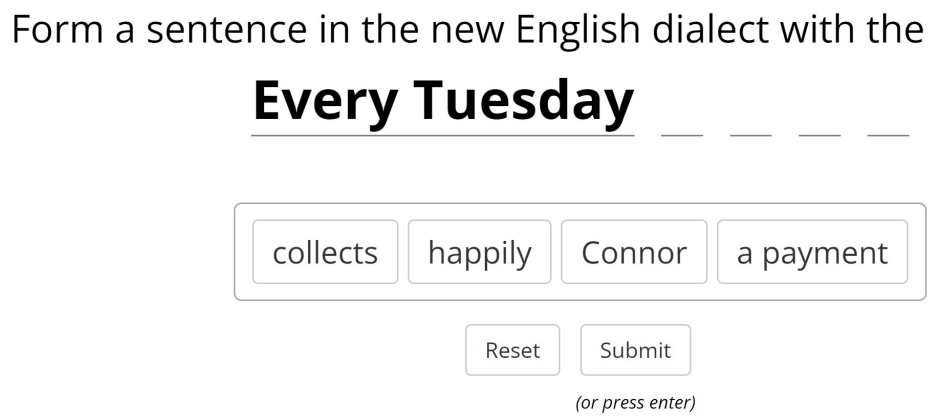


Figure 3.2: Example trial for production task during training. The initial constituent was always shown on the screen, the remaining constituents were provided as buttons beneath. Participants had to simply click on words to construct a sentence. The order of the constituents was randomised for each participant.

sufficient time for reading. After the full sentence was visible, participants were asked to click on one of the constituents, as in Figure 3.1. Feedback was provided after each trial (shown for 450ms if correct; 850ms if incorrect). Participants' performance on this task was used as pre-registered exclusion criteria: Only participants who achieved a score of 90% or higher were included for further analysis (cf. §3.1.1).[6,7] Participants completed 30 trials of this kind.

Participants then moved on to production training (see Hopman & MacDonald 2018 for evidence of the benefit of production during language learning). On each trial, an initial constituent was provided together with four blank lines, as in Figure 3.2. The remaining constituents appeared underneath, each in a separate box. Participants were instructed to fill in the blanks by clicking to insert each constituent into the sentence. All words had to be used. The order of the buttons was randomised for each trial and participant. Participants received feedback after each trial (shown for 1500ms if correct; 3000ms if incorrect). Any incorrectly placed constituents in the sentence produced were highlighted in red. Participants completed 60 trials of this type.

The testing phase was comprised of two parts: sentence production and judgement. Sentence production in testing was identical to production in training, except no initial

[6]Note that this task is mainly to check that participants are attending to the training. It does not require participants to actually learn anything about the language.

[7]A reviewer asks whether the high number of exclusions (cf. §3.1.1) is caused by the difficulty of the attention task, citing the technical terminology (e.g. *passively involved entity* for objects) as a potential reason. We do not think that this is the case. First, participants were given a thorough explanation of this task. Second, conditions were affected differently. In the uniform and adjunct-dominant condition, we had to exclude 23 and 19 participants, respectively. The attrition rate for the object-dominant condition was much higher with 40 participants. This suggests that the issue lies with the difficulty participants have in learning the object-dominant pattern, discussed further below.

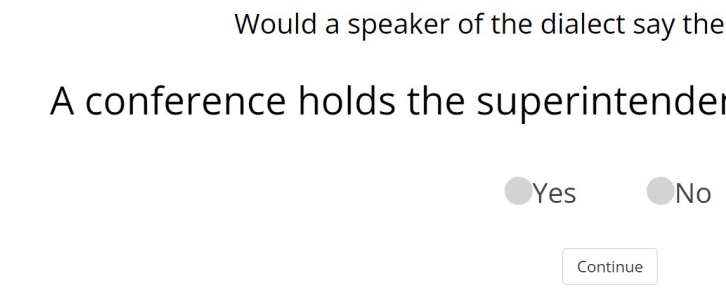


Figure 3.3: Example trial for judgement task during testing. Participants had to decide whether the sentence could be uttered by a speaker of the language.

constituent was provided, and no feedback was given. Participants completed 12 trials of this kind. In the judgement task, participants were asked to judge whether a speaker of the dialect would say a given sentence. On each of the 32 trials, a sentence appeared, and participants chose 'yes' or 'no' as in Figure 3.3. Again no feedback was given. The experiment finished with a questionnaire inquiring participants' language background and strategies employed during the experiment. Table 3.2 provides a summary of the outlined design.

3.2 Predictions

Recall that we operationalised learning of a V2 language as generalising, or extrapolating XP-fronting to novel constituent types: namely indirect objects and complex adjuncts. Our hypothesis was that greater variability in types of clause-initial constituents should aid learning of V2. Accordingly, we made two specific predictions, both measuring the degree to which participants generalise XP-fronting. First, we predicted that participants in the uniform condition would be more likely to produce or accept V2 sentences with initial novel constituent types than participants in the skewed conditions. Second, we predicted that the difference in acceptability between ungrammatical V3 sentences and grammatical V2 sentences with novel constituent types would be greater in the uniform condition compared to the skewed conditions. Recall that neither of these sentence types will have been seen by participants, thus in principle they could both be treated as ungrammatical in the language. However, if participants learn XP-fronting as a generalisable feature, then they should judge fronted novel constituent types as grammatical but V3 sentences as ungrammatical. We predicted learners in the uniform condition should be more likely to do this than learners in the skewed conditions. We made no predictions regarding any differences between the two skewed conditions.

3.3 Results

Our analyses were conducted in R (R Core Team 2020) using the packages lme4 for the statistical analysis (Bates et al. 2015) and ggplot2 for plotting (Wickham 2016). The

Part	Task type	*n* trials	Feedback
Training	Reading	30	yes
	Production	60	yes
Testing	Production	12	no
	Judgement	32	no

Table 3.2: Summary of experimental design.

results of the hypothesis-confirming and exploratory analyses are summarised in Table 3.3.

3.3.1 Hypothesis-confirming analysis

Figure 3.4 shows the proportion of different constituent types in V2 sentences produced by learners in each condition. Recall that indirect objects and complex adjuncts are *novel* constituent types, and all remaining constituents are *familiar* (i.e. seen in initial position during training). Regardless of the type, all constituents use novel lexical items. We first assess whether learners in the uniform condition produced more V2 sentences with the novel constituent types in clause-initial position. We fitted a mixed-effects logistic regression model to all V2 sentences that include a novel constituent type in the sentence production test. The dependent variable was production of a clause-initial novel constituent (=1) or an familiar constituent type (=0). The model included CONDITION (object-dominant, adjunct-dominant or uniform) as a fixed effect, and by-participant and by-item random intercepts. CONDITION was treatment coded with object-dominant as baseline. The model revealed that participants in the adjunct-dominant condition used more novel constituents clause-initially than the participants in the object-dominant condition ($\beta = 3.46$, SE $= .45$, $p = 2.16\times10^{-14}$). Learners in the uniform condition also produced significantly more novel constituents clause-initially than learners in the object-dominant condition ($\beta = .91$, SE $= .45$, $p = .04$). To directly compare the adjunct-dominant and uniform condition, we fitted a model with identical effect structure to the same data with the adjunct-dominant condition as baseline. This model revealed that participants in the adjunct-dominant condition used more novel constituents clause-initially than the participants in the uniform condition ($\beta = -2.55$, SE $= .43$, $p = 2.75\times10^{-9}$). These findings do not straightforwardly match our predictions. First, the skewed conditions unexpectedly differ from one another. Second, while learners in the uniform condition indeed fronted more novel constituents than those in the object-dominant condition, they fronted *fewer* novel constituents than learners in the adjunct-dominant condition.

Turning to the judgement data, we created a new factor SENTENCE TYPE by grouping V2 sentences with initial simple adjuncts and direct objects together as *V2-familiar*.[8] V2

[8]We did not include subject-initial *V2-familiar* sentences as those would be too similar to participants' L1.

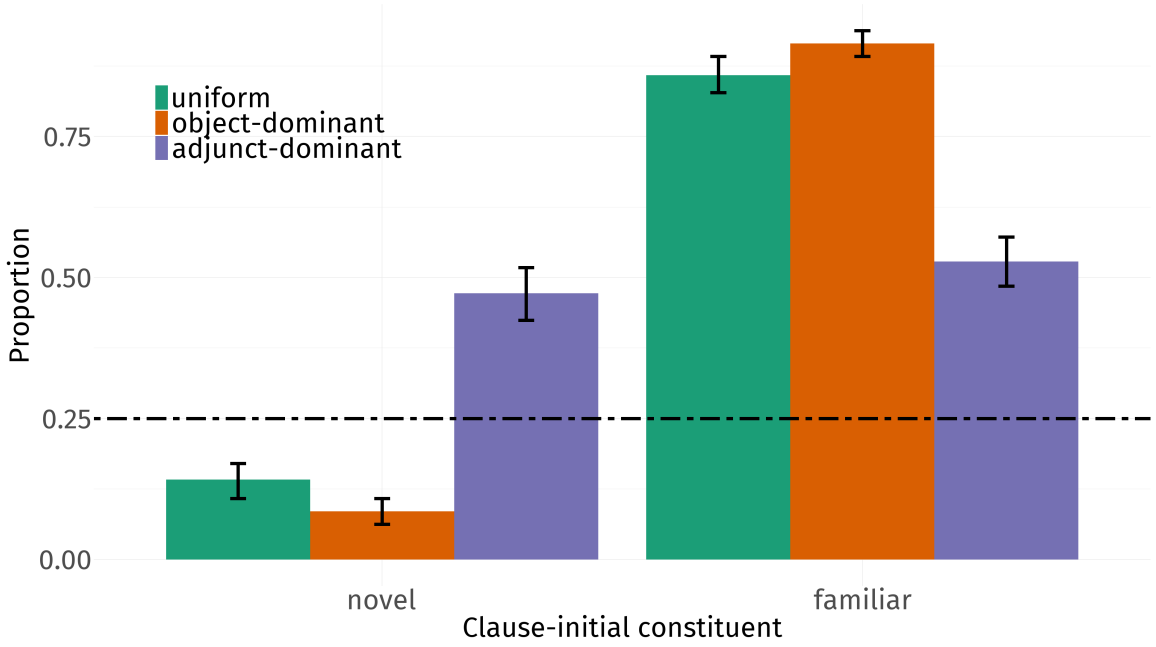


Figure 3.4: Proportion of novel and familiar constituent types in V2 sentences by condition in the sentence production test. The latter are the complement of the former, hence the values of each condition add up to 1. Error bars represent bootstrapped 95% confidence intervals around by-participant means; dotted line indicates chance level. Participants in the adjunct-dominant condition produce significantly more V2 sentences with novel constituents in initial position than the other two conditions. Participants in the uniform condition in turn produce significantly more novel constituent types in the clause-initial position.

sentences with an initial complex adjunct or an indirect object were grouped together as *V2-novel*. All remaining combinations were grouped together as *V3*. The ratings for all three sentence types are shown in Figure 3.5. We then tested whether learners in the uniform condition were (i) more likely to accept V2-novel sentences compared to the skewed conditions and (ii) less likely to accept V3 sentences compared to learners in the skewed conditions. We fitted a mixed-effect logistic regression model to the V2-novel and V3 data. The model included CONDITION and SENTENCE TYPE as fixed effects as well as an interaction term for both. The model also included by-participant and by-item random intercepts and by-participant random slopes for SENTENCE TYPE. Both fixed effects were treatment coded with the object-dominant condition and V2-novel as reference level. To assess the first prediction, we investigated the simple effect of condition. The model indicated that participants in the uniform condition were more likely to accept V2-novel sentences than those in the object-dominant condition, as predicted ($\beta = .99$, SE $= .31$, $p = .001$). In fact, the latter group were

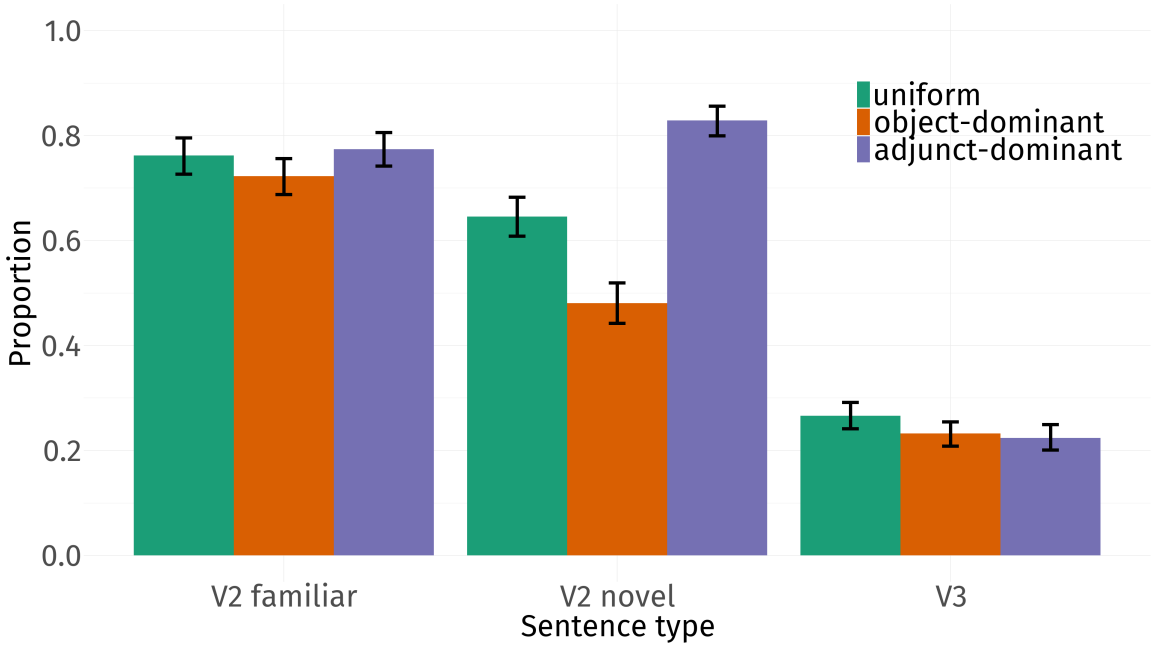


Figure 3.5: Acceptance rates of V2 sentences in the sentence judgement test across three sentence types: V2-familiar, i.e. with clause-initial constituent types seen during training, V2-novel, i.e., with novel clause-initial constituent types, V3, i.e., ungrammatical sentences. Error bars indicate bootstrapped 95% confidence intervals of the mean. V2-novel sentences were accepted at a higher rate in the adjunct-dominant condition compared to other two conditions. V2-novel sentences were also accepted at a higher rate in the uniform condition compared to the object-dominant condition.

not significantly more likely than chance to accept such sentences ($\beta = -0.096$, SE = .25, $p = .70$). The model further showed that V2-novel sentences were significantly more likely to be accepted by learners in the adjunct-dominant condition compared to the object-dominant condition ($\beta = 2.44$, SE = .32, $p = 3.16 \times 10^{-14}$). To compare the adjunct-dominant and uniform conditions, a further model was fitted to the data with identical effect structure but the adjunct-dominant condition as baseline. This model revealed that learners in the uniform condition were significantly less likely to accept V2-novel sentences compared to participants in the adjunct-dominant condition, contrary to our prediction ($\beta = -1.45$, SE = .32, $p = 7.01 \times 10^{-6}$).

To assess the second prediction, i.e., the discrimination of grammatical V2-novel sentences and ungrammatical V3 sentences, we looked at the interaction between CONDITION and SENTENCE TYPE. As Figure 3.5 suggests, V3 clauses were generally less likely to be accepted than V2-novel sentences by learners in the object-dominant condition ($\beta = -1.58$, SE = .36, $p = 1.04 \times 10^{-5}$). We did not find a significant interaction between V3 and the uniform condition ($\beta = -0.82$, SE = .46, $p = .07$)

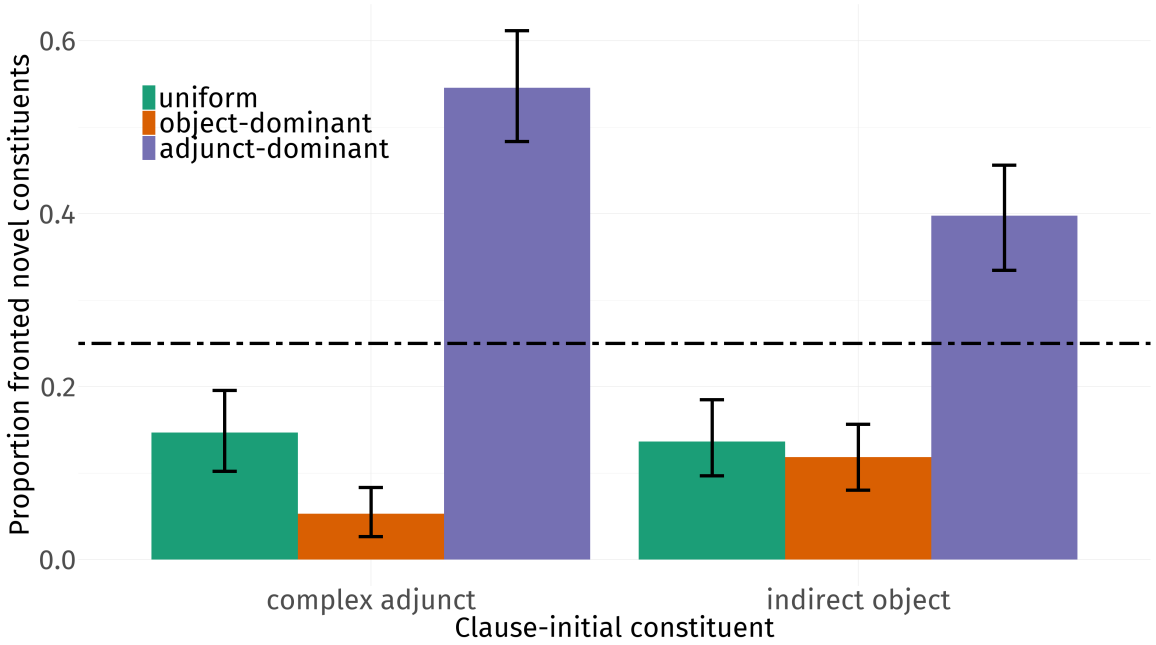


Figure 3.6: Proportion of fronted complex adjuncts and indirect objects in V2 sentences by condition in the sentence production test. Error bars indicate bootstrapped 95% confidence intervals, while the dotted line represents the chance level. Learners in the adjunct-dominant condition use both constituents types with significantly higher proportion in initial position than learners in the other two conditions.

suggesting that the discrimination was not greater in the uniform condition, contrary to our prediction. At the same time, participants in the adjunct-dominant condition were better than those in the object-dominant condition at discriminating V2-novel and V3, indicated by a significant interaction of V3 and the adjunct-dominant condition ($\beta = -2.79$, SE $= .48$, $p = 4.63 \times 10^{-9}$). When learners in the uniform condition were directly compared to those in the adjunct-dominant condition, we again found greater discrimination of V2-novel and V3 for learners in the adjunct-dominant condition ($\beta = 1.95$, SE $= .47$, $p = 2.87 \times 10^{-5}$). Taken together, our second prediction was therefore not borne out: We only found a learning advantage for participants in the adjunct-dominant condition.

3.3.2 Exploratory analysis

Recall that we used indirect objects and complex adjuncts to measure participants' generalisation of V2 in both testing tasks. Arguably, these two types of novel constituents are different from one another. Complex adjuncts are the same type of constituent as simple adjuncts, but longer than any of constituents encountered during training. Indirect objects are a completely novel type of constituent, which participants have not

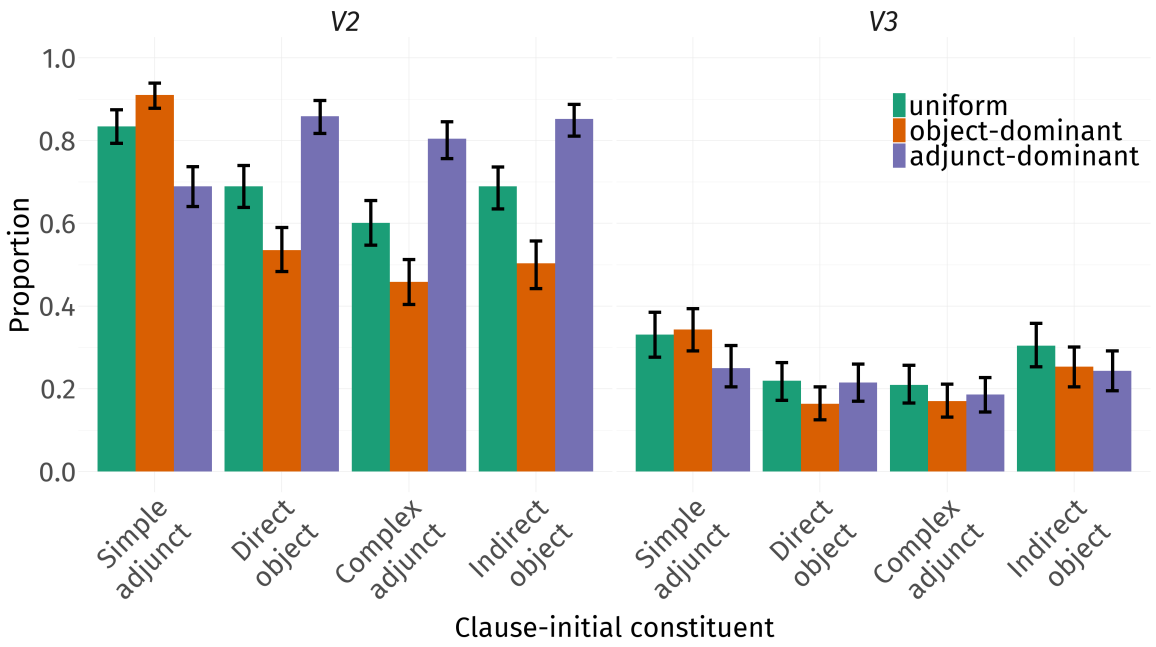


Figure 3.7: Acceptance rate of V2 and V3 sentences with different clause-initial elements by condition in the sentence judgement task. Error bars indicate bootstrapped 95% confidence intervals. Participants in the adjunct-dominant condition are more likely to produce both simple and complex adjuncts in clause-initial position compared to the other two conditions.

encountered in the language at all. It might be that learners in the adjunct-dominant condition exhibited a clear advantage over learners in the other two conditions on the grounds of their familiarity with clause-initial adjuncts in general. If this were the case, we should find a difference between complex adjuncts and indirect objects. As Figure 3.6 shows, participants in the adjunct-dominant condition produce both complex adjuncts and indirect objects more frequently in initial position than participants in the other conditions. This is confirmed in two mixed-effects logistic regression models fitted to V2 sentences that include a complex adjuncts and indirect objects, respectively. The dependent variable was initial constituent type (either complex adjunct or indirect object = 1, other constituents = 0). The model included CONDITION as fixed effect and by-participant and by-item random intercepts. CONDITION was treatment coded with the adjunct-dominant condition as baseline. Participants in the adjunct-dominant condition placed significantly more complex adjuncts and indirect objects in clause-initial position compared to participants in the object-dominant condition (complex adjuncts: $\beta = -5.96$, SE $= .95$, $p = 2.84 \times 10^{-10}$; indirect objects: $\beta = -3.57$, SE $= .82$, $p = 1.41 \times 10^{-5}$) and the uniform condition (complex adjuncts: $\beta = -4.43$, SE $= .92$, $p = 1.72 \times 10^{-6}$; indirect objects: $\beta = -3.18$, SE $= .85$, $p = 1.75 \times 10^{-4}$).

Figure 3.7 depicts the acceptance rate of V2 and V3 sentences with different initial constituents. Similar to the production data, V2 sentences with initial complex adjuncts and indirect objects were both more likely to be accepted in the adjunct-dominant condition compared to the other two. This was confirmed with two mixed-effect logistic regression models fitted to participants' judgements of sentences with clause-initial complex adjuncts and indirect objects, respectively. The model included CONDITION as fixed effect and by-participant and by-item random intercepts. CONDITION was treatment coded with adjunct-dominant as baseline (complex adjuncts: uniform: $\beta = -0.55$, SE $= .21$, $p = .01$; object-dominant: $\beta = -1.17$, SE $= .22$, $p = 6.52\times10^{-8}$; indirect objects: uniform: $\beta = -0.31$, SE $= .22$, $p = .02$; object-dominant: $\beta = -1.02$, SE $= .22$, $p = 3.15\times10^{-6}$). To summarise, neither the production nor judgement data support the idea that the learning advantage observed for participants in the adjunct-dominant condition can be attributed to the similarity between simple and complex adjuncts.

Our hypothesis was specifically about the generalisability of XP-fronting, the placement of verbs within sentences is obviously one of the defining features of V2 too. We thus conducted an exploratory analysis of the verb placement in sentence production. We fitted a mixed-effect logistic regression model to participants' production data. The dependent variable was the position of the verb (second position = 1, not in second position = 0). The model included CONDITION and ADDITIONAL CONSTITUENT (i.e., simple adjunct, complex adjunct and indirect object), as fixed effects as well as their interaction. The latter was included to be sure that verb order was not conditioned on which other constituent was present apart from subject, verb and direct objects. The model also included by-participant random slopes for ADDITIONAL CONSTITUENT. Both fixed effects were treatment coded with object-dominant and simple adjunct as baseline. The model revealed no differences in the likelihood of producing the verb in second position across conditions or additional constituents ($\beta_{min} = -0.62$, $\beta_{max} = 0.14$, $p_{min} = .12$, $p_{max} = .91$).

We conducted a further exploratory analysis, to investigate the acceptance rate of V2-familiar sentences in the sentence judgement test. We did this to check whether the learning advantage for the adjunct-dominant condition is also visible for sentence types that are familiar to participants. Figure 3.5 indicates a generally high acceptance rate for V2-familiar sentences across all three conditions, although the adjunct-dominant condition does show the highest ratings. We fitted a mixed-effect logistic regression model to all V2-familiar sentences with CONDITION as fixed effect and by-participant and by-item random intercepts. CONDITION was again treatment coded with the object-dominant condition as the baseline. The model did not indicate a significant difference between the object-dominant condition and the uniform condition ($\beta = .376$, SE $= .23$, $p = .11$) and the adjunct-dominant condition ($\beta = .31$, SE $= .23$, $p = .19$). This finding suggests that participants in all three conditions learnt the language during training equally well. However, it is worth noting that acceptance ratings do differ to some degree depending on the type of initial constituent, as shown in Figure 3.7. Surprisingly, learners in the object-dominant condition are most likely to accept sentences with clause-initial simple adjuncts, while learners in the adjunct-dominant condition are

	Comparison	Finding
Hypothesis-conf. analysis	Produced novel const. types in initial position in V2 sentences	A-dom > O-dom Uni > O-dom A-dom > Uni
	Acceptance rate V2-novel	Uni > O-dom A-dom > O-dom A-dom > Uni
	Discrimination V2-novel & V3 in judgements	Uni = O-dom A-dom > O-dom A-dom > Uni
Exploratory analysis	Produced complex adjuncts in initial position in V2 sentences	A-dom > O-dom A-dom > Uni
	Produced indirect objects in initial position in V2 sentences	A-dom > O-dom A-dom > Uni
	Acceptance rate V2 sentences with initial complex adjuncts	A-dom > O-dom A-dom > Uni
	Acceptance rate V2 sentences with initial indirect objects	A-dom > O-dom A-dom > Uni
	Produced V2 sentences	no differences
	Acceptance rate V2-familiar sentences	O-dom = A-dom O-dom = Uni

Table 3.3: Summary of the main findings of the hypothesis-confirming and exploratory analyses for the production and judgement data comparing the uniform (uni), adjunct-dominant (A-dom) and object-dominant (O-dom) conditions. '>' indicates a statistically significant contrast, '=' a non-significant contrast.

most likely to accept sentences with clause-initial direct objects. To test this statistically, we fitted a mixed-effects logistic regression model to all V2 sentences with direct objects and simple adjuncts in clause-initial position. As we were interested in the difference between the skewed conditions, the uniform condition were not included in this analysis. The model included fixed effects for CONDITION and INITIAL CONSTITUENT as well as their interaction and by-participant random slopes for INITIAL CONSTITUENT. Both fixed effects were sum-coded with the object-dominant condition and direct objects as reference levels. The model revealed a main effect for INITIAL CONSTITUENT ($\beta = .36$, SE $= .16$, $p = .02$) but not for CONDITION ($\beta = .12$, SE $= .15$, $p = .43$). This suggests the skewed conditions do not differ with respect to their grand means, and that adjuncts were overall rated a bit higher than direct objects. The model further showed a significant interaction between INITIAL CONSTITUENT and CONDITION ($\beta = -1.12$, SE $= .14$, $p = 9.84 \times 10^{-16}$). This confirms our observation that the acceptance rates for these two constituent types differ across conditions. It is unclear why we see this unexpected pattern of results, and it is contradicted by the production data, where,

for example, participants in the object-dominant condition were highly likely to front direct objects (see Figure 3.9).

3.4 Discussion

This experiment investigated whether the distribution of initial constituents in the input impacts learning of V2. Following Yang (2000), we identified non-subject-initial V2 sentences as a crucial type of evidence for V2. Our hypothesis, inspired by evidence for the benefit of variability in other domains (cf. Raviv, Lupyan & Green 2022), was that high variability in initial grammatical functions would aid learners in identifying a key feature of V2, the generalisability of XP-fronting. To test this, we compared three distributions of clause-initial elements: a uniform distribution, an object-dominant skewed distribution, and an adjunct-dominant skewed distribution. We taught participants a semi-artificial language, with English vocabulary but V2 word order. We then asked them to produce and judge sentences with novel constituents in the clause-initial position. These could be either complex adjuncts — longer than any initial constituents seen during training — or indirect objects — a grammatical role not seen in initial position during training. We analysed participants' extrapolation of the clause-initial position to these novel constituent types.

First, it is worth noting that participants in all three conditions were able to learn the requirement that the verb be in second position. This suggests that at least this aspect of V2 is readily learnable in an artificial language. However, the results regarding the generalisability of XP-fronting were mixed. Participants in the uniform condition fronted more novel constituent types in production and were more likely to accept sentences with novel constituent types in clause-initial position compared to participants in the object-dominant condition. However, the apparent advantage of the uniform condition over the object-dominant condition did not extend to the adjunct-dominant condition. Instead, participants in the adjunct-dominant condition were more likely to produce and accept sentences with novel clause-initial constituents compared to both other conditions. Participants in the adjunct-dominant condition were also better at discriminating between grammatical and ungrammatical (V3) sentences. Our exploratory analysis suggested that the learning advantage for the adjunct-dominant condition is not attributable to the similarity between simple and complex adjuncts: participants in the adjunct-dominant condition were also more likely to produce and accept fronted indirect objects. Further exploratory analyses also indicated no overall differences between the three conditions for ratings of V2 sentences with initial familiar constituent types and the frequency of V2 productions. This suggests that a skewed distribution with adjuncts as the dominant element mainly affected generalisation to novel structures.

To summarise, the uniform language gave learners an advantage over the object-dominant condition, but contrary to our prediction, the opposite was the case for the adjunct-dominant condition. Why would there be such a substantial difference between the two skewed conditions? One possibility is that this difference is due to the influence of English on participants' perceptions of sentences in the novel language. This is

particularly relevant here as we used English vocabulary. For example, it may be that for English speakers, object-initial sentences are particularly unexpected. Recall that all objects in the language were inanimate NPs. If participants generally assume that whichever NP is first will be the subject, then object-initial sentences will result in a garden-path effect, or at least a semantic clash, since the inanimate NPs cannot be coerced into subjects. By contrast, the violation induced by adjunct-initial sentences may not be perceived as equally serious; it is a syntactic violation, since the subject is not in the expected position, but it is less likely to induce a garden-path. Further, in English, although displacement of both arguments (7a) and adjuncts (7b) to the clause-initial position is possible, displaced adjuncts are preferred over displaced arguments (Doherty 2005).[9]

(7) a. A block of Emmentaler *the mouse* **found** in the pantry.
b. In the morning, *the mouse* **devours** Appenzeller cheese.

Thus, compared to the adjunct-dominant condition, the object-dominant condition may be more different from participants' native language. This might lead participants in the object-dominant condition to learn the V2 grammar less well, particularly relative to the adjunct-dominant condition, but also to the uniform condition (which still has fewer object-initial sentences). However, it is worth noting that our results suggest that familiar V2 sentences were actually learnt equally well across conditions. It is not entirely clear why a difference in similarity to English would specifically affect generalisation.

It is perhaps worth noting here that English has been described as residual V2 language (e.g Rizzi 1996, Holmberg 2015, Sailor 2020). That is, certain structures still require V2 orders in present-day English such as *wh*-questions (8a) and locative inversions (8b). One might thus expect that these vestiges of a former fully-fledged V2 grammar could constitute another way in which participants' native language affect their performance in the experiment.

(8) a. What kind of cheese **does** the *cheese monger* recommend?
b. Here **is** *your loaf of cheese*.

However, there are good reasons to believe that V2 residues in the English grammar played little to no role in the experiment. Work in L1 acquisition has demonstrated the conservative manner with which learners approach the learning task (Westergaard 2009a). For instance, the Norwegian dialect of Tromsø does not require a strict V2 order in *wh*-questions in that V2 is contingent on the length of the *wh*-word and information-structural aspects (e.g. Westergaard & Vangsnes 2005). Crucially, Westergaard (2009a) showed that learners exhibit target-like structures from early on while no erroneous generalisations to other contexts (such as declaratives) are being made. Similarly for L2/L3/L*n* acquisition, Westergaard (2021) argues for a property-by-property transfer from previously learnt language(s) during the acquisition of a new language. Under

[9]Overall, structures like these are a minority. According to Yang (2000: 242), non-subject-initial sentences account for less than 10% of all cases in the Penn Treebank.

such a model, any influence from the structure of English *wh*-questions on declaratives (as used in the present experiment) is not predicted. Accordingly, locative inversion in English often occurs with unaccusative verbs (e.g, *come*, *sit*) and *be*, that is a well-defined class of verbs (Westergaard 2009b: 68). Generalising this pattern is thus also not expected. Moreover, structures like (8b) are relatively rare (Anderssen & Bentzen 2018: 15) such that any influence would presumably be minimal. Besides, the same question we noted in the previous paragraph arises: Why would the residual V2 grammar affect generalisation in particular?

A second possibility is that there is a genuine learning advantage of having a high proportion of adjunct-initial sentences. In our stimuli, subjects and objects were always DPs, and were not distinguished based on any formal criteria (i.e., case marking). If learners were sensitive to variability in initial constituents defined in terms of grammatical category ([XP-V]) instead of grammatical function ([S|O|A-V]), as we assumed, a high proportion of adjuncts would mean more variability in the input. While Yang (2000, 2002) assumes the latter, the view that grammatical categories are the relevant cue to V2 is indeed taken by Lightfoot (1999, 2006). When the conditions are re-considered from this perspective, learners in the uniform condition saw 33.3% non-DPs and 66.6% DPs clause-initially, learners in the object-dominant condition saw 20% non-DPs and 80% DPs, but learners in the adjunct-dominant condition saw 60% of the initial constituents were non-DPs and 40% DPs. The object-dominant condition thus exhibits the greatest skew and hence the least evidence for generalised XP-fronting, while the adjunct-dominant condition which exhibits the most uniform distribution and thus the most evidence for generalised XP-fronting. This aligns perfectly with our results: generalisation was best in the adjunct-dominant condition followed by the uniform condition, followed by the object-dominant condition. Learning in the adjunct-dominant condition may have been further facilitated by our use of both PPs and AdvPs as adjuncts; this would increase the variability of grammatical categories even more.

It is also possible that both of these explanations for our results are at play: The adjunct-initial condition might facilitate learning the most due to its high level of variability in the grammatical category of initial constituents, while the object-dominant condition might lead to particular poorly learning because it has both low category variability *and* is least similar to English. Future experimental work could tease these apart.[10] Below,

[10]We thank an anonymous reviewer for pointing out a third possible explanation for the advantage of the adjunct-dominant condition over the object-dominant condition: Learners may tend to search for the base word order of a new language using the position of verbal arguments. This is conceivable given the early sensitivity of children to the argument structure of verbs (Naigles 1990, Perkins & Lidz 2021). If such a bias was in place, participants in the object-dominant condition would face a significantly more difficult task in that they need to overcome the native English SVO order in favour of an OVS word order. This is potentially supported by the fact that the majority of all productions exhibit an OV order and hardly any SV orders (cf. Figure 3.9). At the same time, there are two issues with this analysis. On the one hand, the results illustrated in Figure 3.7 are not compatible with this explanation. If learners posit an underlying OVS order, sentences with clause-initial direct objects should exhibit the highest acceptance rate, not sentences with clause-initial adjuncts. On the other hand, it remains unclear what base word order learners in the adjunct-dominant condition would assume as both subjects and objects occur with the same frequency in the clause-initial position. That is the input contains inconclusive evidence as to the base order. One would need to stipulate that

we further examine the distribution of clause-initial elements in natural V2 languages. If adjuncts are indeed beneficial for the acquisition of V2, one may expect to find them over-represented in initial position.

4 The distribution of clause-initial constituents in Germanic

In the previous section, we tested whether greater variability in the distribution of grammatical roles in clause-initial position in the input would lead to better learning of V2. Contrary to our predictions, participants in the adjunct-dominant condition exhibited the best learning outcomes. As discussed above, one possible explanation is that learners benefit from a high frequency of clause-initial adjuncts in the input for the acquisition of V2. Under such an analysis, although there will be many factors that determine the distribution of different elements in initial position, one may expect to find adjuncts over-represented clause-initially in V2 languages precisely because they support learning of V2. The present section will summarise findings in the literature suggesting that adjuncts indeed occur frequently in initial position (§4.1). We will corroborate these earlier findings with the results of a large-scale corpus study of German (§4.2).

4.1 Previous evidence

A number of previous studies have explored the distribution of clause-initial constituents in contemporary Germanic languages.[11] In this section, we will discuss some of these studies, summarised in Table 3.4. An early study on German by Winter (1961), examined 63,000 sentences from diverse sources (theatre, fiction, non-fictional prose, scientific texts) and reported a heavily skewed distribution for written German: Subjects dominate the clause-initial position (66.7%) followed by adverbs and PPs (28.1%). Clause-initial direct objects, on the other hand, are very infrequent, occurring in only 2.9% of all clauses. Crucially, such a skewed distribution appears not to be confined to formal registers: Bohnacker & Rosén (2008) observe a similar distribution in elicited informal texts. Similarly, Engel (1974: 212) provides evidence for a skewed distribution in spoken language, with subjects occurring most in clause-initial position (51.31%), again followed by adjuncts (35.35%) and objects (9.25%).

A skewed distribution of initial constituents is not unique to German and has been noted for other Germanic languages as well. Bohnacker & Rosén 2008 and Bohnacker & Lindgren 2014 show for spoken and written Swedish that, as in German, subjects are the most frequent clause-initial element while direct objects are the least frequent type. Similar patterns have also been observed for spoken Danish (Puggaard 2019)[12] and for

learners transfer the SVO order of their native language in order to explain the observed learning advantage of the adjunct-dominant condition.

[11] Note that there is not a consistent definition of adjuncts in the studies reported here and in §4.3 and results have been reported with different levels of detail.

[12] The authors thank Rasmus Puggaard for bringing this paper to our attention.

Author	Lang	Modality	*n*	S	DO	A	Other
Win(61)	Ger	written	63,000	66.7	2.9	28.1	2.3
Eng(74)	Ger	spoken	5,000	54.31	9.25	35.35	4.09
B&R(08)	Ger	written	1,173	50	7	42	1
B&R(08)	Sw	written	545	71.2	4.4	22.6	1.8
B&L(14)	Sw	written	680	67.0	1.3	32.4	0.3
B&L(14)	Sw	spoken	755	64.4	1.6	33.8	0.3
B&L(14)	Dt	written	646	53.0	2.3	44.7	9.0
B&L(14)	Dt	spoken	711	60.9	1.0	38.1	0.0
Yan(00)	Dt	spoken	n/a	66.8	1.2	23	7
Pug(19)	Dn	spoken	500	62.0	9.4	24.4	4.2

Table 3.4: Proportion of clause-initial subjects (S), direct objects (DO), adjuncts (A) and other elements in German (Ger), Swedish (Sw), Danish (Da) and Dutch (Dt). Authors are abbreviated as follows: Win(61) = Winter 1961, Eng(74) = Engel 1974, B&R(08) = Bohnacker & Rosén 2008, B&L(14) = Bohnacker & Lindgren 2014, Yan(00) = Yang 2000, Pug(19) = Puggaard 2019. Despite differences between but also within languages, all of the listed studies found a skewed distribution of clause-initial elements in different V2 languages. Adjuncts are the second-most frequent element in initial position after subjects. However, note the small sample size of many of these studies.

spoken and written Dutch (Bohnacker & Lindgren 2014). Finally, Yang (2000: 242) found a similar skew even in child-directed speech in Dutch. Despite these similarities, the actual frequencies of different constituent types varies between (but also within) languages. Two observations from Table 3.4 are worth pointing out here: First, adjuncts occur relatively frequently in clause-initial position even though preposed adverbs and PPs occur more frequently in Dutch and German than in other Germanic languages. Second, other frontable constituents such as indirect objects are exceedingly rare in first position (Winter 1961: 201, Puggaard 2019: 298).

Although frequent clause-initial adjuncts appear to be a cross-Germanic feature, a cautious interpretation of the previous studies is still warranted. All studies (with the notable exception of Winter (1961)) relied on a relatively small sample. Here, we conducted a large-scale corpus study on the distribution of clause-initial elements in present-day German aimed at replicating the general findings reported in previous smaller studies.

4.2 Evidence from a large-scale corpus study on German

To test the validity of previous findings, we conducted a corpus study using a significantly larger sample.[13] We chose the dependency treebank TüBa-D/DP (de Kok &

[13]The experimental code as well as the analysis is available online.

Corpus	*n* sentences
wiki	45.5 M
europarl	2.2 M
speeches	619,152

Table 3.5: Number of sentences in the three examined subcorpora of TüBa-D/DP, i.e. Wikipedia (wiki), proceedings of the European Parliament (europarl) and speeches of German officials (speeches) (de Kok & Pütz 2019: 1).

Pütz 2019) as all sentences are annotated for their position in the topological field (cf. Drach 1937, Wöllstein 2010). This annotation enables automatic identification of V2 and non-V2 clauses as well as clause-initial constituents. TüBa-D/DP consists of four different subcorpora: Wikipedia articles (wiki), proceedings of the European Parliament (europarl), speeches of German officials (speeches) and newspaper articles from *taz*. The first three are publicly available and were used as the basis for the present study. While the three corpora differ in size, as illustrated in Table 3.5, even the smallest of the three (speeches) is still significantly larger than any of the corpora in §4.1.

We built a custom python script to count where and with what frequency each syntactic function (as annotated in the corpora) occurs relative to the finite verb. Only sentences ending in a full stop or colon were taken into consideration as the word order in interrogative as well as imperative and exclamative sentences may differ. Furthermore, sentences with fewer than two words were excluded. Crucially, each clause in a sentence was considered separately given that multiple V2 clauses are grammatical.

Due to the different sizes of the three corpora, distributions were plotted separately for each corpus in Figure 3.8.[14] However, all three corpora illustrate a similar pattern: non-clausal subjects are the dominating clause-initial element followed by adjuncts, while direct objects occur only very infrequently in initial position. Adverbial clauses (which were considered distinct from adjuncts due to their clausal status) are more frequently attested in initial position in europarl and wiki corpus than direct objects.[15] All other types of constituents that the grammar permits in initial position are very rare. This is particularly striking in the case of indirect objects, which are perfectly grammatical in initial position, and yet almost never occur there in these corpora.

[14] Attentive readers may have noticed that the proportions do not add up to 1 but rather to a value between approximately 0.96 and 0.98. A likely explanation for this divergence lies in errors in the part-of-speech tagging.

[15] Note that this is a somewhat artificial distinction as adverbial clauses are formally adjunct CPs. However, we wanted to underscore the fact that they are comparable in frequency with direct objects.

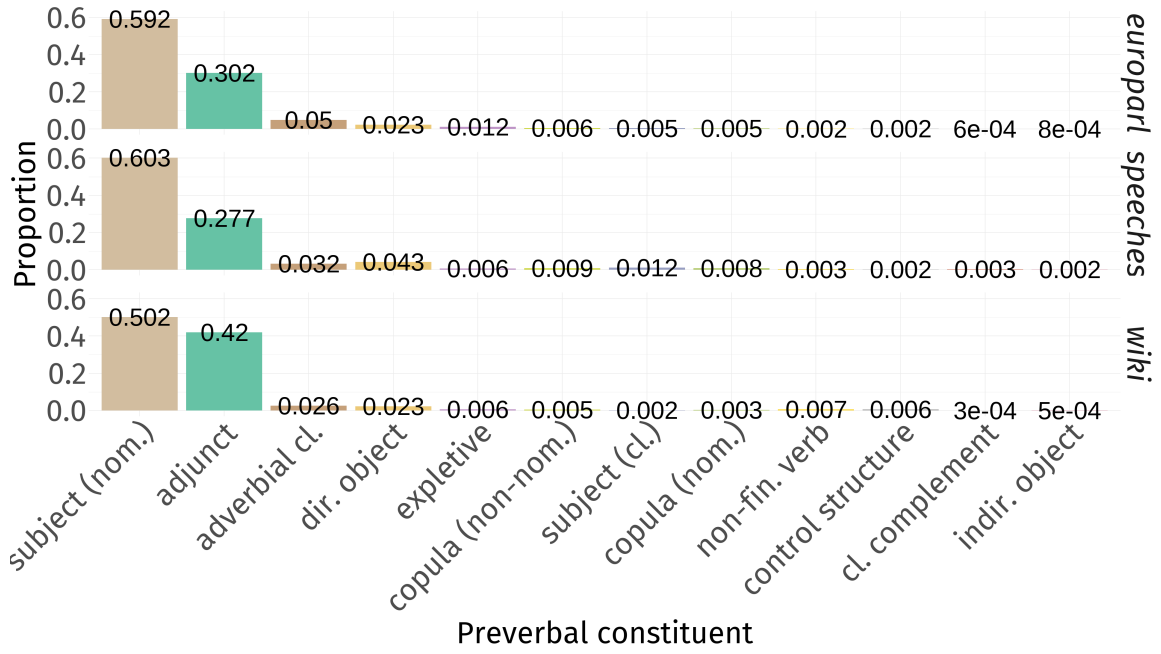


Figure 3.8: Proportion of different constituents in clause-initial position in V2 clauses in the Proceedings of the European Parliament (europarl), speeches of German officials (speeches) and Wikipedia (wiki). All three corpora show a skewed distribution similar to the ones previously observed in the literature (cf. §4.1). Note the parallels between the two spoken corpora europarl and speeches in terms of the frequencies of subjects, adjuncts and direct objects.

To summarise, our corpus study confirms earlier findings of a skewed distribution of clause-initial constituents in V2 sentences.[16] To summarise, there is robust evidence that adjuncts appear very frequently in initial position in the Germanic languages.

4.3 Comparison with Old Romance

The way in which different constituent types are distributed in the clause-initial position has also been systematically studied for the Medieval Romance languages (e.g. Labelle & Hirschbühler 2018: 281, Wolfe 2018: 25). Here, there is also evidence that adjuncts were strikingly frequent. Consider the proportion of different types of constituents in clause-initial position according to Wolfe (2018: 25) summarised in Table 3.6. The

[16] We also examined whether the skew persists when frequencies of clause-initial elements proportional to their base rates are considered. Intuitively, one could imagine that the base rate of subjects occurring in a sentence is greater than the base rate of adjuncts, and thus the former have a greater probability of appearing in initial position simply due to this. However, Monte Carlo simulations revealed that constituents are still skewed, with subjects and adverbial clauses more likely to be fronted given their base rate. In the europarl and wiki corpus, adjuncts also appear more frequently in clause-initial position than expected.

Language	Subj	Obj	Adj
Old French	46.32	12.84	40.84
Old Occitan	23.78	25.00	51.22
Old Sicilian	47.63	12.62	39.75
Old Venetian	74.34	9.26	16.40
Old Sardinian	62.26	18.24	19.50
Old Spanish	35.32	6.42	58.26

Table 3.6: Distribution of different clause-initial constituents in V2 sentences in several Old Romance languages (Wolfe 2018: 25). Note that Wolfe (2018) makes a more fine-grained distinction for adjuncts that are subsumed under the label adjunct here. All of the listed languages show a substantial amount of adjunct-initial sentences. In Old Venetian and Old Sardinian the proportion of initial adjuncts is lower but still larger than the proportion of initial objects.

frequency of adjuncts in these languages was comparable, if not higher, to the ones found for modern Germanic languages (cf. Table 3.4). Importantly, in later stages, at least for French, when V2 is being lost, a different picture emerges. Steiner (2014: 129), for instance, notes an increase of SV structures from the 13th (47.11%) to the 16th century (62.88%) in V2 sentences.[17] Simultaneously, the frequency of null subjects remained mostly constant (approx. 30%) in the same contexts. These data support the possibility that the decline of V2 grammar is connected to the decline in adjunct-initial sentences.

In English, which has also lost V2, non-subject-initial constructions like (9) are possible.

(9) Here **is** *the platter with gruyère cheese*.

However as noted above, there is evidence that these are extremely low frequency (Anderssen & Bentzen 2018: 15).[18] More generally, in English, it is very uncommon for any element other than the subject to be in initial position (less than 10% of sentences

[17] Adjunct-initial and object-initial sentences are not separately considered by Steiner (2014). However, assuming a lower proportion of object-initial sentences is justified given prevalence of this pattern across V2 language.

[18] Roeper (1999: 175) however reports incidental evidence that children generalise this pattern to some extent for a very brief period:

(i) what **calls** *that*
'What is it called'
(Roeper 1999: 175)

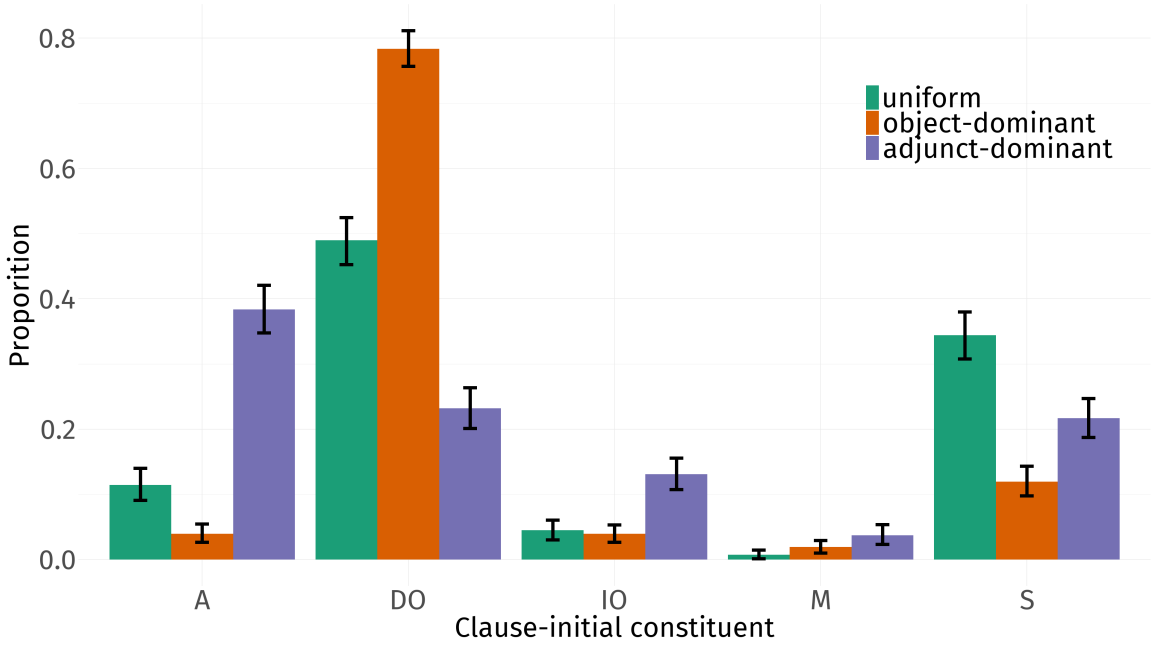


Figure 3.9: Proportion of different clause-initial constituent types across different trial types in V2 sentences by condition in the sentence production test of the experiment (cf. §3). A = adjuncts, DO = direct objects, IO = indirect objects, M = adverbials to mark verbal movement, S = subjects. Learners in the object-dominant condition produced a high number of object-initial sentences. This exceeds their input significantly, which contained only 60% of clause-initial objects.

according to Yang 2000: 242).[19] This is in line with the idea that there is something special about having adjuncts frequently in initial position for the maintenance of V2.

5 General discussion

V2 is a striking feature of almost all Germanic languages, and yet it is cross-linguistically rare, and its loss has been well-documented, including in English, Welsh and almost all Romance language. The loss of V2 has been tied to a lack of sufficient evidence in learners' input, crucially a lack of non-subject-initial V2 sentences (Lightfoot 1999, 2006, Yang 2000, 2002, Willis 1998, Meelen 2016). Combined with domain-general evidence for the benefit of variability on learning (cf. Raviv, Lupyan & Green 2022), we hypothesised that a V2 language will be learnt best if the clause-initial position is occupied by a maximally diverse set of grammatical functions, i.e. subjects, direct objects

[19] See also Westergaard, Lohndal & Lundquist (2023) for evidence of the relationship between the production of non-subject-initial declaratives and V2 errors in heritage speakers of Norwegian living in the USA.

and adjuncts. We conducted an ALL experiment to test this hypothesis. We compared three conditions, a uniform condition in which subjects, direct objects and adjuncts occurred with identical frequency in the clause-initial position. Additionally, we ran two skewed conditions — one skewed towards direct objects and one towards adjuncts. Contrary to our prediction, learning was best in the adjunct-dominant condition followed by the uniform condition. Two possible explanations were identified. First, it could be that object-initial sentences are particular odd for our native English-speaking participants, and the condition which has the least such sentences — the adjunct-dominant condition — is the best learnt. Second, there could be a genuine advantage of having adjuncts in clause-initial position with a high frequency. Although further work is necessary to distinguish between the two explanations, the second is supported by the fact that our adjunct-initial condition actually had the most evidence for variability in terms of grammatical categories (rather than roles or functions). Our results therefore provide some degree of support for the view held by Lightfoot (1999, 2006) who points to changes in the distribution of clause-initial categories as an explanation for the loss of V2. In our case, the adjunct-initial condition featured a more balanced mixed of DPs, PPs, and AdvPs, while the object-initial condition featured an over-representation of DPs in initial position. The former led to better generalisation of XP-fronting, and the latter to diminished generalisation of XP-fronting. Interestingly, in this condition, participants' productions were even more skewed than their input, as can be seen in Figure 3.9. Such a heavily skewed distribution in productions was not observed for any of the other conditions. In fact, participants in the adjunct-dominant condition produced the least skewed distribution. If the output of learners in the object-dominant condition served as the input for a next generation of learners, then the evidence for V2 in this generation would be even more reduced (Kirby, Cornish & Smith 2008). This constitutes a plausible trajectory for the loss of V2, paralleling the situation e.g. in French where the evidence for V2 was more and more reduced (Roberts 1993, Steiner 2014).

If adjuncts are important more generally in the learning of V2, and not just in our experiment, one might expect to see a significant number of clause-initial adjuncts in exactly those Germanic languages that have retained a V2 grammar. To test this claim, we reviewed work on the distribution of clause-initial constituents in Germanic languages. We also reported a large-scale corpus study on German. This analysis confirmed that while subjects occur most frequently in the initial position, the next most common constituent type is adjuncts. By contrast, in languages which are in the process of losing V2, there is evidence for a particular reduction in frequency of adjuncts in initial position. In languages, like English, which no longer have V2, adjuncts are very unlikely to occur in initial position. These distributional differences between natural languages support the idea that adjuncts may play a special role in the learning and maintenance of V2.

6 Conclusion

In this paper, we investigated the hypothesis that learning of a V2 language is fostered by a maximally variable distribution of clause-initial constituents in the input. This hypothesis was derived from work suggesting that a V2 grammar will be lost if learners do not receive sufficient evidence for the grammar (Lightfoot 1999, 2006, Yang 2000, 2002), and that variability in general is good for learning (Gómez 2002). We tested this by comparing learning of an artificial V2 language with subjects, objects and adjuncts in clause-initial position with equal frequency, with learning of languages with skewed frequencies — either object-dominant or adjunct-dominant. While our results suggest that different distributions of clause-initial elements do indeed affect learning outcomes, learners were best able to generalise XP-fronting to novel constituent types when the distribution of initial elements was skewed towards adjuncts. They were least able to generalise when the distribution was skewed towards objects. We propose that a high frequency of adjuncts in initial position is in fact likely to be an important feature of V2. It may lead to higher variability in the grammatical categories of elements in first position, which could be more important than the variation in grammatical roles (Lightfoot 1999, 2006). Further, there is robust evidence, including from our large-scale corpus study of German, showing that adjuncts are highly frequent in initial-position in current (or historical) V2 languages, but not in a non-V2 language like English. Our results therefore support the idea that diminished evidence in the input can lead to the loss of V2. Our study also adds to the body of literature now demonstrating the utility of artificial language learning studies in understanding language typology and change.

Bibliography

Anderssen, Merete & Kristine Bentzen. 2018. Different outcomes in the acquisition of residual V2 and do-support in three Norwegian-English bilinguals: Cross-linguistic influence, dominance and structural ambiguity. *Frontiers in Psychology* 9. 1–18. doi:10.3389/fpsyg.2018.02130.

Axel, Katrin. 2007. *Studies on Old High German syntax. Left sentence periphery, verb placement and verb-second*. Amsterdam, Philadelphia: John Benjamins. doi:10.1075/la.112.

Axel, Katrin. 2009. The consolidation of verb-second in Old High German: What role did subject pronouns play? In Paola Crisma & Giuseppe Longobardi (eds.), *Historical syntax and linguistic theory*, 128–143. Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780199560547.003.0008.

Axel-Tober, Katrin. 2018. Origins of verb-second in Old High German. In Agnes Jäger, Gisella Ferraresi & Helmut Weiß (eds.), *Clause structure and word order in the history of German*, 22–47. Oxford: Oxford University Press. doi:10.1093/oso/9780198813545.003.0003.

Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1–48. doi:10.18637/jss.v067.i01.

Benincà, Paola. 1995. Complement clitics in medieval Romance: The Tobler-Mussafia law. In Adrian Battye & Ian Roberts (eds.), *Clause structure and language change*, 325–344. New York, Oxford: Oxford University Press. doi:10.1093/oso/9780195086324.003.0012.

Bohnacker, Ute & Josefin Lindgren. 2014. Fundament, formellt subjekt och frekvens: Ordföljdsmönster i svenska, nederländska och hos vuxna inlärare av svenska. *Språk & Stil* 24. 33–71.

Bohnacker, Ute & Christina Rosén. 2008. The clause-initial position in L2 German declaratives. Transfer of information structure. *Studies in Second Language Acquisition* 30(4). 511–538. doi:10.1017/S0272263108080741.

Cournane, Ailís & Espen Klævik-Pettersen. 2023. The role of the conservative learner in the rise and fall of verb-second. *Journal of Historical Syntax* 7(9). 1–48. doi:10.18148/hs/2023.v7i6-19.160.

Culbertson, Jennifer. 2012. Typological universals as reflections of biased learning: Evidence from artificial language learning. *Language and Linguistics Compass* 6(5). 310–329. doi:10.1002/lnc3.338.

Culbertson, Jennifer. 2017. New approaches to Greenbergian word order dependencies. In N. J. Enfield (ed.), *Dependencies in language: On the causal ontology of linguistic systems*, 23–38. Berlin: Language Science Press. doi:10.5281/zenodo.573778.

Culbertson, Jennifer. 2023. Artificial language learning. In Jon Sprouse (ed.), *The Oxford handbook of experimental syntax*, 271–300. Oxford: Oxford University Press. doi:10.1093/oxfordhb/9780198797722.013.9.

Culbertson, Jennifer & David Adger. 2014. Language learners privilege structured meaning over surface frequency. *Proceedings of the National Academy of Sciences* 111(16). 5842–5847. doi:10.1073/pnas.1320525111.

Culbertson, Jennifer & Kathryn Schuler. 2019. Artificial language learning in children. *Annual Review of Linguistics* 5. 353–373. doi:10.1146/annurev-linguistics-011718-012329.

Doherty, Monika. 2005. Topic-worthiness in German and English. *Linguistics* 43(1). 181–206. doi:10.1515/ling.2005.43.1.181.

Drach, Erich. 1937. *Grundgedanken der deutschen Satzlehre*. Frankfurt am Main: Diesterweg.

Engel, Ulrich. 1974. Syntaktische Besonderheiten der deutschen Alltagssprache. In Hugo Moser (ed.), *Gesprochene Sprache. Jahrbuch 1972*, 199–228. Düsseldorf: Schwann.

Ettlinger, Marc, Kara Morgan-Short, Mandy Faretta-Stutenberg & Patrick C. M. Wong. 2016. The relationship between artificial and second language learning. *Cognitive Science* 40(4). 822–847. doi:10.1111/cogs.12257.

Eythórsson, Thórhallur. 1995. *Verbal syntax in the early Germanic languages*. Ithaca, NY: Cornell University dissertation.

Fischer, Olga, Ans van Kemenade, Willem Koopman & Wim van der Wurff. 2001. *The syntax of Early English*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511612312.

Getz, Heidi. 2018. *Sentence first, arguments after: Mechanisms of morphosyntax acquisition*. Washington, DC: Georgetown University dissertation. doi:10.31237/osf.io/ske5q.

Getz, Heidi. 2019. Acquisition of morphosyntax: A pattern learning approach. In David W. Lightfoot & Jonathan Havenhill (eds.), *Variable properties in language. Their nature and acquisition*, 103–114. Washington, DC: Georgetown University Press.
Gómez, Rebecca & Jessica Maye. 2005. The developmental trajectory of nonadjacent dependency learning. *Infancy* 7(2). 183–206. doi:10.1207/s15327078in0702_4.
Gómez, Rebecca L. 2002. Variability and detection of invariant structure. *Psychological Science* 13(5). 431–436. doi:10.1111/1467-9280.00476.
Holmberg, Anders. 2015. Verb second. In Tibor Kiss & Artemis Alexiadou (eds.), *Syntax – Theory and analysis. An international handbook*, 342–383. Berlin, München, Boston: De Gruyter Mouton. doi:10.1515/9783110377408.342.
Hopman, Elise W.M. & Maryellen C. MacDonald. 2018. Production practice during language learning improves comprehension. *Psychological Science* 29(6). 961–971. doi:10.1177/0956797618754486.
van Kemenade, Ans. 1987. *Syntactic case and morphological case in the history of English*. Dordrecht: Foris. doi:10.1515/9783110882308.
Kirby, Simon, Hannah Cornish & Kenny Smith. 2008. Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences* 105(31). 10681–10686. doi:10.1073/pnas.0707835105.
Kirby, Simon, Monica Tamariz, Hannah Cornish & Kenny Smith. 2015. Compression and communication in the cultural evolution of linguistic structure. *Cognition* 141. 87–102. doi:10.1016/j.cognition.2015.03.016.
de Kok, Daniël & Sebastian Pütz. 2019. *TüBa-D/DP stylebook*. University of Tübingen: Seminar für Sprachwissenschaft.
Labelle, Marie & Paul Hirschbühler. 2018. Topic and focus in Old French V1 and V2 structures. *Canadian Journal of Linguistics* 63(2). 264–287. doi:10.1017/cnj.2017.52.
de Leeuw, Joshua R., Rebecca A. Gilbert & Björn Luchterhandt. 2023. jsPsych: Enabling an open-source collaborative ecosystem of behavioral experiments. *Journal of Open Source Software* 8(85). 1–4. doi:10.21105/joss.05351.
Lightfoot, David. 1999. *The development of language. Acquisition, change, and evolution*. Malden, MA: Blackwell.
Lightfoot, David. 2006. *How new languages emerge*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511616204.
Martin, Alexander, Annie Holtz, Klaus Abels, David Adger & Jennifer Culbertson. 2020. Experimental evidence for the influence of structure and meaning on linear order in the noun phrase. *Glossa: a journal of general linguistics* 5(1). 1–21. doi:10.5334/gjgl.1085.
Martin, Alexander, Theeraporn Ratitamkul, Klaus Abels, David Adger & Jennifer Culbertson. 2019. Cross-linguistic evidence for cognitive universals in the noun phrase. *Linguistics Vanguard* 5(1). 1–13. doi:10.1515/lingvan-2018-0072.
Meelen, Marieke. 2016. *Why Jesus and Job spoke bad Welsh. The origin and distribution of V2 orders in Middle Welsh*. Leiden: Universiteit Leiden Proefschrift.
Naigles, Letitia. 1990. Children use syntax to learn verb meanings. *Journal of Child Language* 17(2). 357–374. doi:10.1017/S0305000900013817.

Perkins, Laurel & Jeffrey Lidz. 2021. Eighteen-month-old infants represent nonlocal syntactic dependencies. *Proceedings of the National Academy of Sciences* 118(41). 1–7. doi:10.1073/pnas.2026469118.

Puggaard, Rasmus. 2019. Flexibility of frequent clause openers in talk-in-interaction: Det 'it, that' and så 'then' in the prefield in Danish. *Nordic Journal of Linguistics* 42(3). 291–327. doi:10.1017/S0332586519000088.

R Core Team. 2020. *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

Raviv, Limor, Gary Lupyan & Shawn C. Green. 2022. How variability shapes learning and generalization. *Trends in Cognitive Sciences* 26(6). 462–483. doi:10.1016/j.tics.2022.03.007.

Rebuschat, Patrick & John N. Williams. 2012. Implicit and explicit knowledge in second language acquisition. *Applied Psycholinguistics* 33(4). 829–856. doi:10.1017/S0142716411000580.

Rizzi, Luigi. 1996. Residual verb second and the wh-criterion. In Adriana Belletti & Luigi Rizzi (eds.), *Parameters and functional heads. Essays in comparative syntax*, 63–90. New York, Oxford: Oxford University Press. doi:10.1093/oso/9780195087932.003.0003.

Roberts, Ian. 1993. *Verbs and diachronic syntax. A comparative history of English and French*. Dordrecht, Boston, London: Kluwer. doi:10.1007/978-94-011-2910-7.

Roberts, Ian. 1996. Remarks on the Old English C-system and the diachrony of V2. In Ellen Brandner & Gisella Ferraresi (eds.), *Language change and generative grammar*, 154–167. Opladen: Westdeutscher Verlag. doi:10.1007/978-3-322-90776-9_6.

Roeper, Thomas. 1999. Universal bilingualism. *Bilingualism: Language and Cognition* 2(3). 169–186. doi:10.1017/S1366728999000310.

Ruiz, Simón, Kaitlyn M. Tagarelli & Patrick Rebuschat. 2018. Simultaneous acquisition of words and syntax: Effects of exposure condition and declarative memory. *Frontiers in Psychology* 9. 1–11. doi:10.3389/fpsyg.2018.01168.

Sailor, Craig. 2020. Rethinking 'residual' verb second. In Rebecca Woods & Sam Wolfe (eds.), *Rethinking verb second*, 126–149. Oxford: Oxford University Press. doi:10.1093/oso/9780198844303.003.0006.

Steiner, Brittany Devan Jelm. 2014. *The evolution of information structure and verb second in the history of French*. Bloomington: Indiana University dissertation.

Tagarelli, Kaitlyn M., Simón Ruiz, José Luis Moreno Vega & Patrick Rebuschat. 2016. Variability in second language learning. *Studies in Second Language Acquisition* 38(2). 293–316. doi:10.1017/S0272263116000036.

Vikner, Sten. 1995. *Verb movement and expletive subjects in the Germanic languages*. New York, Oxford: Oxford University Press. doi:10.1093/oso/9780195083934.001.0001.

Waldmann, Christian. 2008. *Input och output. Ordföljd i svenska barns huvudsatser och bisatser*. Lund: Lunds Universitet Dissertation.

Westergaard, Marit. 2009a. *The acquisition of word order. Micro-cues, information structure, and economy*. Amsterdam, Philadelphia: John Benjamins. doi:10.1075/la.145.

Westergaard, Marit. 2009b. Word order in Old and Middle English: The role of information structure and first language acquisition. *Diachronica* 26(1). 65–102. doi:10.1075/dia.26.1.03wes.

Westergaard, Marit. 2021. Microvariation in multilingual situations: The importance of property-by-property acquisition. *Second Language Research* 37(3). 379–407. doi:10.1177/0267658319884116.
Westergaard, Marit, Terje Lohndal & Björn Lundquist. 2023. Variable V2 in Norwegian heritage language. An effect of crosslinguistic influence? *Linguistic Approaches to Bilingualism* 13(2). 133–162. doi:10.1075/lab.20076.wes.
Westergaard, Marit R. & Øystein A. Vangsnes. 2005. Wh-questions, V2, and the left periphery of three Norwegian dialect types. *Journal of Comparative Germanic Linguistics* 8(1). 117–158. doi:10.1007/s10828-004-0292-1.
Wickham, Hadley. 2016. *ggplot2. Elegant graphics for data analysis*. Cham: Springer. doi:10.1007/978-3-319-24277-4.
Willis, David. 1998. *Syntactic change in Welsh. A study on the loss of verb-second*. Oxford: Clarendon. doi:10.1093/oso/9780198237594.001.0001.
Winter, Werner. 1961. Relative Häufigkeit syntaktischer Erscheinungen als Mittel zur Abgrenzung von Stilarten. *Phonetica* 7(4). 193–216. doi:10.1159/000258110.
Wolfe, Sam. 2018. *Verb second in Medieval Romance*. Oxford: Oxford University Press. doi:10.1093/oso/9780198804673.001.0001.
Wöllstein, Angelika. 2010. *Topologisches Satzmodell*. Heidelberg: Winter.
Yang, Charles. 2000. Internal and external forces in language change. *Language Variation and Change* 12(3). 231–250. doi:10.1017/S0954394500123014.
Yang, Charles. 2002. *Knowledge and learning in natural language*. New York: Oxford University Press. doi:10.1093/oso/9780199254149.001.0001.
Yang, Charles. 2016. *The price of linguistic productivity*. Cambridge, MA: MIT Press. doi:10.7551/mitpress/9780262035323.001.0001.
Þorgeirsson, Haukur. 2012. Late placement of the finite verb in Old Norse Fornyrðislag meter. *Journal of Germanic Linguistics* 24(3). 233–269. doi:10.1017/S1470542712000037.

3.3 Conclusion

The present chapter reported the results of the third experiment in this thesis. The experiment tested the hypothesis that less variability in the clause-initial position in terms of the attested grammatical functions (e.g. subject, object) should affect the learning of a V2 language negatively. Three conditions were compared: a uniform condition in which subjects, direct objects as well as adjuncts were realised with the same frequency in clause-initial position, and two skewed-conditions where either direct objects or adjuncts dominated the clause-initial position. The results only partially confirmed my hypothesis. As predicted, participants in the uniform condition showed better learning (i.e. generalisation) than participants in the object-dominant condition. Participants in the adjunct-dominant condition, however, showed better learning than the other two conditions. Two explanations for the observed patterns were proposed. On the one hand, participants' native language might have impacted their learning such that participants in the object-dominant condition were negatively affected compared to those in the adjunct-dominant condition. This could be attributable to processing or the fact that topicalisations in English are more common with adjuncts than with objects. On the other hand, participants in the adjunct-dominant condition might have had a genuine learning advantage. Such an advantage can readily be explained if participants are not sensitive to grammatical functions but rather to grammatical *categories* (e.g. NP, PP) when learning a V2 grammar. When the distributions of different clause-initial grammatical categories are examined, it becomes evident that learners in the adjunct-dominant condition were exposed to the most uniform language. Participants in the object-dominant condition in turn were trained on the most skewed language in terms of grammatical categories. The results then directly follow from the hypothesis that high variability in the clause-initial grammatical categories benefits learning: The language with the highest variability (adjunct-dominant condition) was learnt best, while the language with the least variability (object-dominant condition) was learnt worst. Recall from Chapter 1 that the focus on grammatical functions as evidence was a mere stipulation — there was no motivation other than the fact that grammars are usually formulated in terms of grammatical functions. That is, nothing should conflict with such a change. The experimental results of the subsequent chapter, in fact, will provide support for learners' sensitivity to variability in clause-initial grammatical functions and grammatical categories.

The results of the large-scale corpus study that was also presented in the current chapter further support the second explanation for the experimental findings. Although subjects are the most frequent clause-initial constituent type in German, adjuncts constitute the second most frequent one. What is more, the results from Monte Carlo simulations briefly mentioned in footnote 16 show that this distribution can still be observed when the frequencies of clause-initial elements relative to their base rates are considered.[20] These findings also converge with earlier studies on the distribution of clause-initial elements in Germanic and Old Romance V2 languages. Crucially however, languages that are in the process of losing their V2 grammar, show a decline in clause-

[20]The results of the simulation are reported in more detail in Appendix B.

initial adjuncts, as already noted in §1.3. The distribution in natural languages thus supports the idea that adjuncts play a central role for the acquisition of a V2 grammar. It is worth noting that additional corpus analysis in future work could in fact focus on evidence for the relevance of variability in grammatical categories. While the corpus study reported here was conducted before we initially entertained this potential interpretation of the experimental results, it is very likely that more adjuncts will imply more non-DPs in clause-initial position. For example, even if some are DPs (e.g. *next Tuesday* in (10)), the set of adjuncts will certainly include both PPs and AdvPs in addition.

(10) *This pack of cheese* **will** expire next Tuesday.

Two comments on the experimental results are also in order. First, the measurement for the extrapolation of the V2 rule is potentially open to criticism. Recall that indirect objects and complex adjuncts (i.e. adjuncts with three instead of two words) were employed as measure. Arguably, complex adjuncts constitute a weak test of extrapolation given participants' familiarity with (simple) adjuncts in initial position. Even though these constituents comprised a more complex structure than those encountered during training, they would only be genuinely novel if participants exhibited sensitivity towards the number of words in a single constituent. Given that grammatical rules are generally not considered to be defined in terms of the number of words (e.g. 'prepose the third word in every clause'), it is possible that participants did not experience complex adjuncts as novel, or at least not as novel as indirect objects. Although this may cast doubt on whether participants had to extrapolate in sentences featuring complex adjuncts, the extrapolation with indirect objects still holds. Indirect objects formed a completely novel constituent type. The exploratory analysis (§3.3.2) crucially showed that learners in the adjunct-dominant condition performed significantly better at test for sentences with indirect objects than the other two conditions (both in the production and judgement task). That is, the overall pattern of the results still holds when only the strict measure is taken into account.

The second comment pertains to the composition of the clause-initial position in the training stimuli. In the main body of the present chapter, it was claimed that the distribution of clause-initial constituents was the most uniform in the adjunct-dominant condition when grammatical categories instead of functions are considered. This claim was supported by the distribution of DPs and non-DPs in initial position. Those numbers, however, were in fact merely an approximation; some of the adjuncts were actually realised as DPs (e.g. *every day*). I therefore examined the distribution of clause-initial constituents in the training sentences in more detail. For this purpose, all training items were coded for the syntactic category of the clause-initial constituent. Proper nouns were coded as DPs, adverbial phrases and prepositional phrases as AdvPs and PPs, respectively. Table 3.7 summarises the distribution of the syntactic categories across conditions. Overall, the same pattern emerges as previously claimed in §3.4: Participants in the adjunct-dominant condition were exposed to the input exhibiting the highest variability, whereas the training items in the object-dominant condition were the least variable when it comes to clause-initial grammatical categories. This can be seen in

Condition	DP-initial	PP-initial	AdvP-initial
Uniform	.73	.22	.05
Object-dominant	.84	.13	.03
Adjunct-dominant	.52	.40	.08

Table 3.7: Distribution of clause-initial syntactic categories in the training stimuli of experiment 3. While the distribution for initial DPs, PPs and AdvPs was identical for all participants in the uniform condition, the proportions of the object-dominant and adjunct-dominant conditions varied for individual participants. That is, the reported values are the means in the respective conditions. The variation between participants in the same condition was the result of the random selection of repeated sentences (cf. §3.1.2).

the lower proportion of clause-initial DPs in the adjunct-dominant condition compared to the other conditions. Crucially however, DPs still accounted for the majority of clause-initial constituents in all three conditions. This contrasts with the previous claim for the adjunct-dominant condition that non-DPs exceed DPs in frequency. As Table 3.7 illustrates, the relation between DPs and non-DPs is closer to 50:50 when PPs and AdvP are combined. When it comes to the distribution of AdvPs and PPs, it has to be noted that their distribution was also skewed in that the former occurred significantly more frequently than the latter in all three conditions. The difference was particularly pronounced in the adjunct-dominant condition.[21]

Even though the results of this additional analysis would suggest that DPs and non-DPs were in an equilibrium (in other words they are uniformly distributed) in the adjunct-dominant condition, such a conclusion is somewhat misleading. Subsuming AdvPs and PPs under the same label obscures the high frequency of AdvPs. Furthermore, a measure is need to quantify the actual differences between conditions. A purely descriptive analysis does not suffice, especially when the distribution of the uniform and object-dominant condition is compared. The differences in frequency are very minor for PPs and even for DPs and AdvPs, the differences are small. It is therefore necessary to quantify the variability more reliably. A standard measure for variability is entropy (Shannon 1948). The entropy of the training distribution for each participant across conditions is given in Figure 3.10. While participants in the uniform condition were exposed to the exact same variability during training, the entropy varies between participants in the object-dominant and adjunct-dominant conditions. To determine whether the difference in entropy was actually significant, a linear regression model was fitted to the entropy values of the training stimuli seen by each participant. The model included CONDITION (object-dominant, adjunct-dominant or uniform) as fixed effect. CONDITION was treatment coded with the uniform condition as baseline. The model

[21] The last observation does not come as a surprise considering the higher proportion of AdvPs in the training data — a high proportion of adjuncts in the adjunct-dominant condition causes AdvP to be overly represented clause-initially.

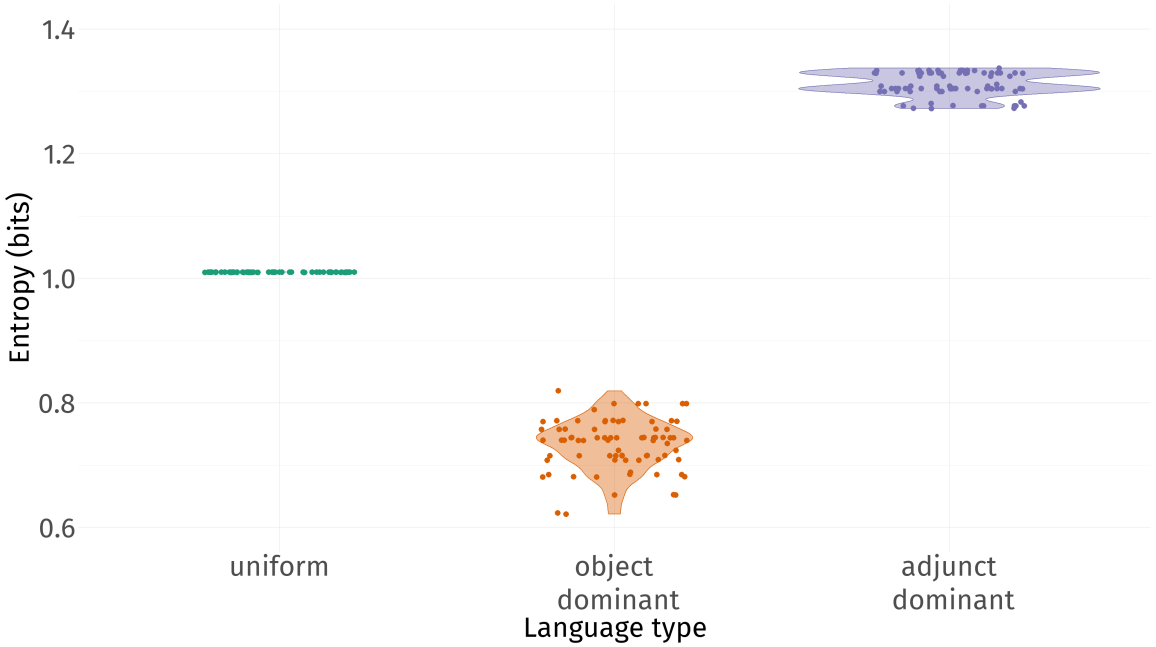


Figure 3.10: Entropy (in bits) in the training in experiment 3 by condition. Coloured dots indicate the entropy experienced by participants in the three conditions. The entropy is highest in the adjunct-dominant condition and the lowest in the object-dominant condition.

indicated that the entropy of the training items was significantly lower in the object-dominant condition than in the uniform condition ($\beta = -0.28$, SE $= .004$, $p < 0.001$). At the same time, the entropy of the adjunct-dominant condition was significantly higher than the one participants experienced in the uniform condition ($\beta = .30$, SE $= .004$, $p < 0.001$). These results thus lend further support to my previous conclusions. The variability in clause-initial grammatical categories is highest in the adjunct-dominant condition and lowest in the object-dominant condition. The interpretation from the main body of this paper can thus be maintained.

Despite the evidence indicating a special status of adjuncts in the acquisition of V2 grammars, it still remains a possibility that participants' L1 (i.e. English) is responsible for the observed patterns. In addition, both explanations might have conspired as they are not mutually exclusive. It is therefore necessary to disentangle the contribution of each factor. A straightforward approach is to alter the nature of the artificial language. If the close resemblance of the semi-artificial language used in the experiment to participants' L1 promoted (adjunct-dominant condition) or hindered learning (object-dominant condition), such effects should be weakened if the artificial language is rendered more dissimilar to participants' L1. One of the arguments for using a semi-artificial language, as argued in §1.6, is the elimination of the lexical learning burden; participants can instead concentrate on learning the syntactic rules of the language. As

Martin et al. (2020) have shown, however, the results obtained with a semi-artificial language by Culbertson & Adger (2014) can be replicated with a fully-artificial language. I thus conducted two further experiments with different types of artificial languages. The experiments will be described in the next chapter. Their results will support both proposed explanations for experiment 3. Moreover, it will be shown that variability in grammatical functions and grammatical categories can foster the learning of a V2 language.

CHAPTER 4

DISTINGUISHING THE EFFECTS OF TRANSFER AND LEARNING ADVANTAGE

4.1 Introduction

The results of experiment 3 have demonstrated that variability in the clause-initial position affects the acquisition of a V2 grammar. Different distributions led to different learning outcomes. Contrary to the initial prediction however, participants who were exposed to a language with the highest variability in the initial position did not display the best learning outcome. Instead, learners of a language in which adjuncts account for the majority of clause-initial elements exhibited the best learning performance. Interestingly, learners of the high variability language still performed better at test than learners of a language where objects constitute the most frequent clause-initial constituent. Two mutually non-exclusive interpretations of these findings were offered in Chapter 3. First, participants' native language (i.e. English) may have contributed to the observed patterns. This would not come as a surprise considering the fact that the artificial language used in experiment 3 used English vocabulary items. Learners do not constitute a tabula rasa, as noted in §1.6. Hence, interference (or transfer) from previously acquired languages is expected (cf. Westergaard 2021b), especially when the same lexicon is used. According to the second interpretation, a high proportion of adjuncts in clause-initial position may confer a genuine learning advantage to learners. This could be readily explained if learners are not sensitive to variability in grammatical functions (e.g. subjects, objects) — as hypothesised in this thesis so far — but rather to variability in grammatical *categories* (e.g. NP, PP). The language in which adjuncts dominate the initial position does exhibit the highest variability of grammatical categories of the three languages tested. The second interpretation would thus maintain a modified version of the hypothesis developed in Chapter 1. Indeed, the focus on variability in grammatical functions was grounded in the convention of formulating grammars in terms of the relative order of grammatical functions, rather than a theoretically motivated choice.

What is more, the results of a large-scale corpus study, also reported in Chapter 3, provide additional evidence for the second interpretation: Although subjects are the most frequent constituent type in clause-initial position, adjuncts still occur with

considerable frequency in clause-initial position in German. This pattern is also mirrored in the distributions in the initial position of Old Romance V2 languages and other modern Germanic V2 languages. If a significant proportion of adjuncts in initial position is necessary for the acquisition and retention of a V2 grammar, the presence of a V2 grammar in these languages follows as a direct results from the observed distributions.

Despite the arguments in favour of the second interpretation — i.e. the fostering effect of frequent clause-initial adjuncts — the results of experiment 3 do not allow me to draw a distinction between the two interpretations. Besides, they are, as argued above, not mutually exclusive; it is conceivable that both factors were at play. The goal of the present chapter is to examine the contribution of participants' L1 to the patterns found in the results of the previous experiment. One way to accomplish this is by reducing the impact of participants' native language. If the effects observed in the original experiment are replicated in a context that involves less influence of the L1, this would support the role of adjuncts (or variation in grammatical categories) in learning V2. In past work, two avenues have been pursued to explore or reduce the effect of participants' L1 in artificial language learning (ALL) experiments. First, researchers have sought to vary the population and explore whether the same behavioural effects hold (Martin et al. 2019). For example, one could target languages that differ with respect to their word orders. Second, researchers have moved to make the language itself less similar to participants' L1. For example, some experiments use a fully-artificial language, with phonological features that are distinct from participants' L1 (Martin et al. 2020). Alternatively, some research has instantiated linguistic patterns in unusual modalities, including visually-based languages (Shapiro & Steinert-Threlkeld 2023), or gesture (Culbertson, Schouwstra & Kirby 2020). In some cases, these changes have replicated earlier findings using a semi-artificial language, suggesting behavioural effects were not driven by L1 influence (Culbertson & Adger 2014). In the set of the present experiments, I am already using speakers of a language that does not have V2, in order to reduce the possibility that experience with this type of language, and a particular distribution of initial elements will confound my interpretation of the findings. In principle, I could test speakers of a language with object before subject, to see whether even for these participants the adjunct language is best. However, such languages are rare (and most speakers bilingual in a subject-initial language). Instead, in this chapter, I focus on the second option. In the following, I first elaborate on the advantage of using a less L1-like artificial language (§4.2). Then, I report the results of two experiments replicating experiment 3; first using a fully-artificial language (§4.3), second using a visual language (§4.4). The results of both experiments will then be discussed (§4.5), before the chapter concludes with a summary of the main findings (§4.6).

4.2 Replicating the effects of experiments with different lexicon types

The goal of this chapter is to test whether a language with a uniform distribution of clause-initial constituent types produces a learning advantage, e.g., over a language with a skew favouring adjuncts in initial position, when the language is less similar to participants' L1. If the adjunct-initial advantage persists, then interference from participants' L1 is not likely to be solely responsible. This would suggest the possibility that having a large proportion of adjuncts in initial position led to the learning advantage, though future work might be needed to conclusively show this. In experiments 1–3, a semi-artificial language was used to ease the burden of lexical learning. This procedure should enable participants to focus on the syntactic patterns rather than lexical learning. Furthermore, this approach generally allows a greater variability of lexical items to be used. More variable lexical items allow for better control of any confounding effects of specific lexical items. In the case of V2, it also increases the type variability of elements occurring in the initial-position. Both these aspects were desirable. However, given that the results suggest the possibility of L1 influence, it makes sense to change this aspect of the design.

4.3 Experiment 4

To test the effect a changed artificial language has, in Experiment 4, I conducted a fully-artificial language learning experiment. Getz (2018) has already demonstrated that a fully-artificial V2 grammar can in principle be learnt under experimental conditions. Issues with the learnability of the language are therefore not expected. Apart from the nature of the language, one further change was made to the experimental design. Recall that no subject-dominant condition was included in experiment 3 on the grounds that the language in this condition would resemble natural English too closely. This restriction does not necessarily hold to the same degree once I move to a fully-artificial lexicon: The relation between the language used in experiment 4 and English is much weaker. Consequently, four instead of three conditions were included: the uniform condition (where subjects, direct objects and adjuncts are realised with the same frequency clause-initially), and three skewed conditions with either subjects, objects or adjuncts as the most frequent element.

I test the same hypothesis as in my previous experiments: More variability in the clause-initial position should lead to better learning of a V2 grammar. For now, the hypothesis is still framed in terms of variability in the grammatical functions (i.e. subjects, direct objects, and adjuncts). For the operationalisation of what it means to learn a V2 language, the same measures as in experiments 1 & 2 were adopted, namely the extrapolation of the verb placement rule and the flexibility of the clause-initial constituent.

4.3.1 Methods

Similar to all previous experiments in this thesis, the present experiment was preregistered before data collection commenced. The ethics board of the Linguistics and English Language department of The University of Edinburgh granted ethics approval (404-2122/2). The experiment was again implemented with the JavaScript library jsPysch (de Leeuw, Gilbert & Luchterhandt 2023).

4.3.1.1 Participants

211 participants were recruited online after the experiment was advertised on Prolific. Participants were paid £3.96 as compensation for their participation. Prolific filters were used to restrict the participant pool to participants from traditionally monolingual English-speaking countries, viz. the US, Canada, UK, Ireland, Australia and New Zealand.[1] As in the previously reported experiments, additional filters were set so that only English monolinguals growing up with English as their earliest and only language could participate. Additionally, English had to be their primary language. Participants' subjects at university could not include English language, English literature or languages more broadly. Finally, an approval rating of 95% or higher was required to be eligible for participation. 14 participants had to be excluded, because of their low performance during training following the preregistered exclusion criterion (cf. §4.3.1.3). No participant indicated any knowledge of a V2 language in a post-test questionnaire. In the analysis of the data, 50 participants in the uniform and object-dominant condition, 49 participants in the adjunct-dominant condition, and 48 participants in the subject-dominant condition were included.

4.3.1.2 Materials

A fully-artificial language was used to construct the stimuli materials. The change from a semi-artificial language to a fully-artificial language necessitated several changes to the materials compared to previous experiments. First, the lexicon size had to be reduced to make the language learnable in a relatively short time frame. Table 4.1 provides the full list of the lexical items and morphological markers used in the experiment. Three major constituent types were included: nominal phrases, prepositional phrases and verbs. NPs can be further subdivided into animate nouns and inanimate nouns. While the former category included animals, the latter category comprised objects. The second type, i.e. PPs, described environmental conditions. All PPs are headed by the same preposition, namely *en*. The final category are verbs. Two different types of verbs were included: three transitive verbs denoting irreversible actions and one ditransitive verb. The inclusion of irreversible transitive actions should ensure that participants could easily identify the actor and patient of a sentence. The language also included a determiner *i* as well as accusative and dative case marking. In order to ease lexical learning, all lexical items were constructed in such a way that they either stand in an onomatopoetic relation with the denoted concept (e.g. *muh* 'cow') or roughly resemble

[1] In contrast to the previous experiments, participants were recruited from a broader group to create a more representative sample.

Category	Subcategory	Lexical items
Noun	animate	*kit* ‘cat’, *muh* ‘cow’, *gak* ‘chicken’, *onk* ‘pig’
	inanimate	*bil* ‘ball’, *tic* ‘tic’, *schin* ‘light bulb’, *bup* ‘trumpet’
PP		*en nih* ‘in the night’, *en driz* ‘in the rain’, *en flek* ‘in the snow’, *en sul* ‘in the sun’
Verb	transitive	*benga* ‘(to) hammer’, *mawa* ‘(to) kiss’, *pifa* ‘(to) shoot’
	ditransitive	*hada* ‘(to) give’
Determiner		*i* ‘the’
Case marker	accusative	*-pi*
	dative	*-ki*

Table 4.1: Lexicon of artificial language used in experiment 4.

the English translation for a related word (e.g. *en flek* ‘in the snow’ from *snowflake*). This design choice enabled the inclusion of more lexical items, which is potentially important, given my hypothesis that variability is key.

Before training sentences were constructed, one animate noun, one inanimate noun as well as one PP were randomly selected for each participant to be withheld during training. The ditransitive verb was also not included in the training materials. The remainder of the words were combined with all transitive verbs to create a total of 81 unique sets that encompassed one element of each constituent type. For each set, a subject-initial, an object-initial and an adjunct-initial sentence was created, following the word order blueprints illustrated in (1a) to (1c). Animate nouns were subjects, inanimate nouns were direct objects and PPs were adjuncts. In each permutation, a determiner was added to the subject and the direct object to form DPs. Furthermore, direct objects were marked by the accusative suffix *-pi*. Note that in contrast to experiments 1–3, no movement marker was included. The meaning of sentence-medial adverbs like *often* or *usually* are difficult to teach in a short experiment. Moreover, movement markers were included in experiments 1–3 to highlight the deviating patterns in the semi-artificial language from participants’ native language. Due to the fully-artificial nature of the language, pointing out the differences did not have the same importance.

(1) a. S-V-DO-A
b. DO-V-S-A
c. A-V-S-DO

From these 243 sentences, 45 were selected randomly for each participant as training sentences. This approach helps to counterbalance any confounding effects arising from specific lexical items. On a par with previous experiments, the distribution of different clause-initial elements was determined by the condition. In the *uniform* condition, subjects, direct objects and adjuncts occurred with the same frequency in clause-initial position. In the *subject-dominant* condition, subjects accounted for the majority of all

Condition	Subject-initial	Object-initial	Adjunct-initial
Uniform	15	15	15
Subject-dominant	27	9	9
Object-dominant	9	27	9
Adjunct-dominant	9	9	27

Table 4.2: Distribution of subject-initial, object-initial and adjunct-initial sentences in each of the four conditions in experiment 4.

clause-initial elements, while in the *object-dominant* condition objects are the most frequent clause-initial element. Accordingly in the *adjunct-dominant* condition, adjuncts dominated the clause-initial position. The dominant element in each of the skewed conditions accounted for 60% of all clause-initial elements. The two non-dominating elements made up 20% each of initial constituent types. In the uniform condition on the other hand, subjects, direct objects and adjuncts were equally distributed — i.e. 33.3%. Based on these distributions, the appropriate amount of subject-initial, object-initial and adjunct-initial sentences was sampled from the set of all sentences. See Table 4.2 for the exact numbers. One consequence of the random sampling was that some participants might have encountered all three sentence versions (i.e. subject-initial, object-initial and adjunct-initial) constructed from the same set, whereas others might have been exposed to either only one or two variants. The sentences were divided into two blocks, while maintaining the distribution of clause-initial elements. The first block comprised 15 sentences, while the larger second block contained 30 sentences. The first block was used for reading trials, the second for production trials (cf. §4.3.1.3).

For the testing phase, two different sets of materials were created. The first set was used for production testing and consisted of 23 unordered sets of constituents. Two subcategories can be distinguished based on their composition: one subcategory contained exclusively familiar constituent types, i.e. subject, direct object, adjunct and verb. The second subcategory included familiar types (i.e. subject, direct object, and verb) but crucially also one novel constituent type, namely an indirect object. Indirect objects were marked by the dative case marker *-ki* (cf. Table 4.1). Similar to the previous experiments, trials with indirect objects allow me to measure how well participants generalise XP-fronting. No complex adjuncts were included, as the lexicon did not comprise any words that could function as modifiers and additional words would be difficult to teach in the short time of the experiment. For both subcategories of testing items, the lexical novelty of individual constituents was manipulated. Recall that one animate noun, one inanimate noun and one PP was withheld during training. These words thus constitute novel lexical items for participants during testing. Using both familiar and novel lexical items allows me to control for potential preferences of participants due to the lexical familiarity of particular items. It is conceivable that participants are more reluctant to place lexically novel words in clause-initial position than lexically familiar words. For the sets with transitive verbs, all possible permutations

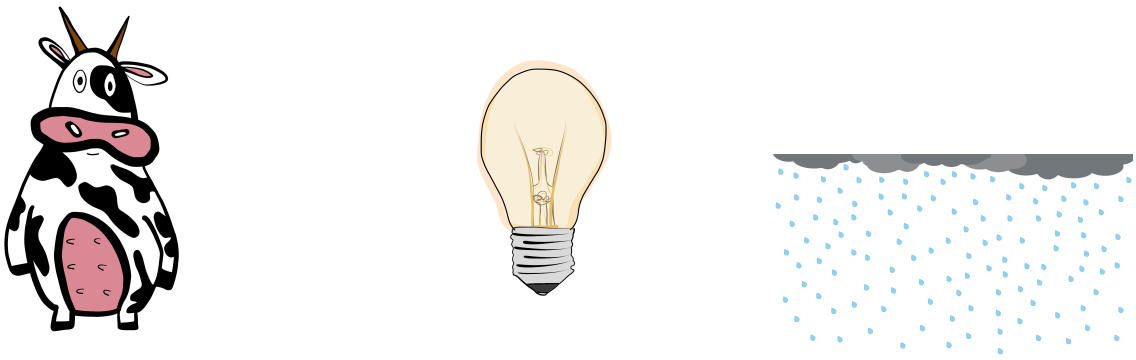
of lexically familiar and novel animate nouns, inanimate nouns and PPs were constructed. The only exceptions were exclusively familiar lexical items, seven different combinations thus remained. One permutation per combination was randomly selected for each participant. That is, seven unordered testing sets with exclusively familiar grammatical functions were included in the testing materials.

For sets with indirect objects, the procedure was slightly altered: First, subjects and indirect objects were never both lexically novel. The semantics of the ditransitive verb *give* dictates that the indirect object (i.e. the recipient) is animate. Given that only one lexically novel animate noun existed, the same noun could not simultaneously function as subject and indirect object. Consequently, either the subject or the indirect object was represented by a lexically novel item, while the other had to be represented by a lexically familiar item. Second, permutations with lexically familiar subjects, direct objects and indirect objects were included, because the ditransitive verb was lexically novel for participants. That is, one constituent was still unfamiliar to participants. For each of the six possible combinations, one permutation of lexical items was randomly selected with the exception of the one where the indirect object was represented by a novel lexical item: The combinations lexically familiar subject, lexically novel direct object and lexically novel indirect object allow three different versions of which all were included. Accordingly for the combinations with familiar subject, familiar direct object and unfamiliar indirect object, nine different options are possible, and all were included for each participant. These testing trials were therefore identical across participants. That is a total of 16 unordered sets with indirect objects were included.

The second set of testing items was used for a judgement task (cf. §4.3.1.3). Sentences were created by crossing the three factors VERB POSITION (V2 or V3), INITIAL CONSTITUENT (subject, adjunct, direct object, or indirect object) and LEXICAL ITEM (lexically familiar initial element or lexically novel initial element). The first two factors are almost identical to the ones in experiments 1–3 — the only difference is that complex adjuncts were replaced by subjects in the present experiment. The third factor on the other hand is new: It encodes whether the clause-initial element was seen during training (= lexically familiar) or withheld during training (= lexically novel). One sentence each was created for level combinations involving familiar clause-initial grammatical functions (i.e. subjects, direct objects and adjuncts). Three sentences for each level combination were created with the initial constituent as an indirect object. That is, a total of 24 test sentences were constructed. Test sentences with V3 order all featured subjects in second position (i.e. XP-S-V) with the obvious exception of subject-initial sentences. In these cases, the adjunct followed the subject (i.e. S-PP-V-DO). On a par with other training and test items for the experiment, judgement sentences were randomly selected for each participant.

4.3.1.3 Procedure

The experiment could be accessed through a web browser on participants' personal computers or laptops. At the beginning of the experiment, participants were informed that the purpose of the experiment is to learn an alien language with the help of materials compiled by a teacher. The experiment was divided into three phases: training



(a) cow (b) light bulb (c) in the rain

Figure 4.1: Three examples for lexical learning in experiment 4.

and testing of lexical items, word order training and word order testing. In the first phase, bare nouns and PPs were introduced; verbs were not separately introduced. One item from each category was withheld from participants to be used as novel lexical item in the last phase of the experiment (cf. §4.3.1.2). First, one image at a time visualising a noun (either animate and inanimate) was shown.[2] Underneath each image, a box was displayed showing the corresponding word in the alien language. After 500ms, a continue button was enabled and participants could proceed to the next trial. Each noun was shown twice in a randomised order. Once participants completed all noun training trials, they were tested on how well they have learnt individual lexical items. In a given trial, participants saw an image representing a noun and two answer options below the image — a foil and the correct answer. After selecting one of the answer options, participants received feedback on their performance. If the correct word was chosen, feedback was shown for 800ms. However, if participants selected the incorrect word, feedback was shown for 1200ms where participants were informed which the correct answer was. Each noun was tested once. The noun training and testing was succeeded by training and testing of PPs. The procedure was identical to the one for nouns. Figure 4.1 provides examples for the images used in the lexical learning part.

In the next phase of the experiment, participants were trained on the word order of the language. This was the first time that participants encountered the verbs. No special instructions were provided for the verbs as their meaning could be deduced from the context. Following the insights from Chapter 2, a reading task and a production task were used to teach participants the word order. At the start of the reading task, participants were informed that they would see sentences in the alien language along with a scene depicting the sentence. Furthermore, they were instructed to identify either the doer (i.e. subject), the doee (i.e. direct object), the action (i.e. verb) or the environment of the action (i.e. adjunct). These terms deviate considerably from the terminology of previous experiments, in that they are less technical than before. Even

[2] I thank Clem Ashton and Kenny Smith for providing me the images used in experiment 4.

Figure 4.2: Example of a reading trial during training in experiment 4. Constituents were successively revealed, one at a time: *i ticpi*, *pifa*, *i kit en flek*. To facilitate learning, the meaning of the sentence was illustrated with an image. Once the whole sentence was visible on the screen, participants had to identify the prompted constituent type by clicking on it.

though the high error rate in the reading task in Chapter 3 was attributed to difficulties with the language and not the task (cf. fn.7 in §3.2), the terminology was nevertheless simplified. During a reading trial, the image illustrating the scene was shown first for 500ms. The orientation of the images was determined randomly: It was either shown normally or vertically flipped. The random orientation was used to prevent any bias relating the position in the sentence to the position of the entity in the image. For instance, without the random orientation, subjects would always appear on the left and (English-speaking) participants might therefore want to produce them first. Sentences were incrementally revealed (constituent-by-constituent) with a delay of 500ms between constituents. As before, the purpose of delaying the revelation of constituents was to provide participants sufficient time for reading. Once the whole sentence was visible on the screen, participants were prompted to select one of the constituents by clicking on it, as shown in Figure 4.2. If participants clicked on the correct constituent, the feedback was shown for 1000ms. If they clicked on the wrong constituent, feedback was shown for 3000ms. Akin to previous experiments, the pre-registered exclusion

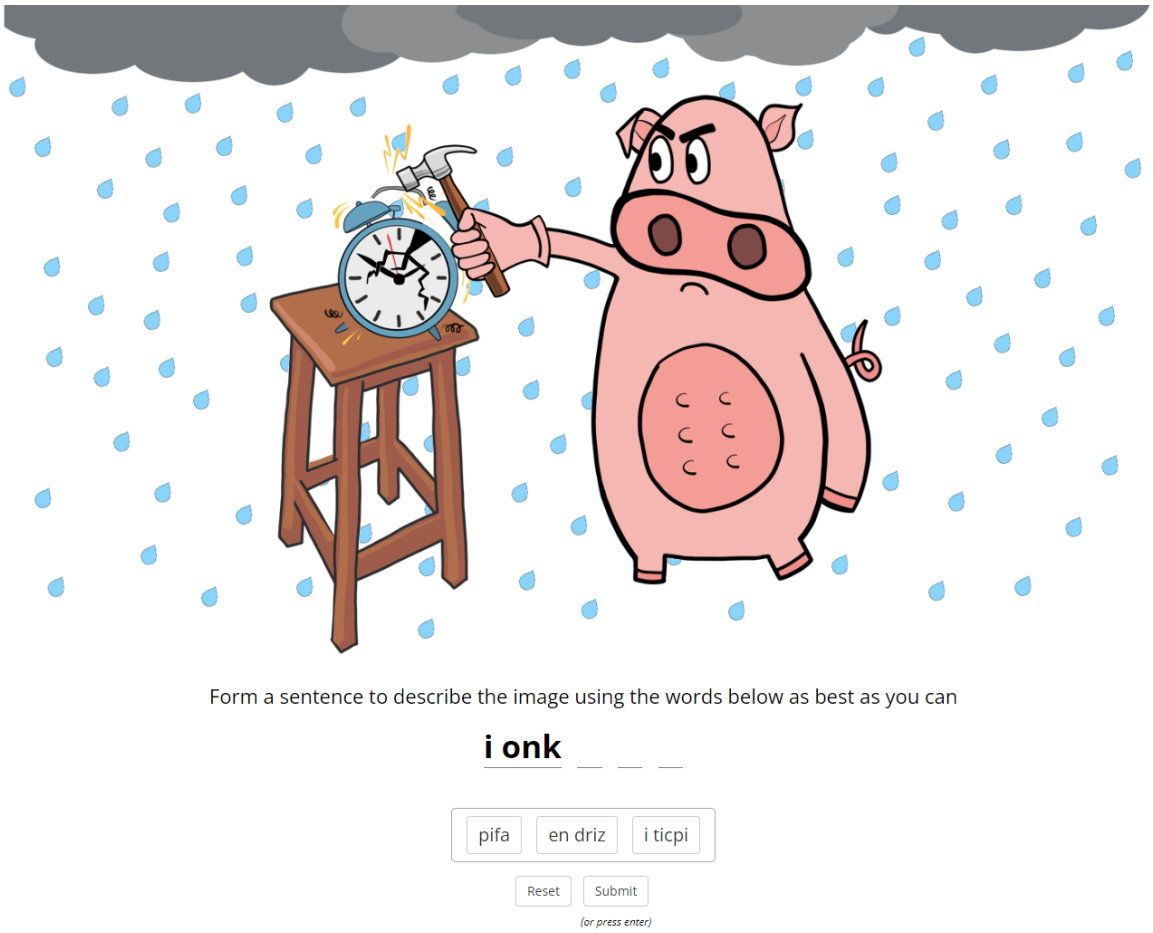


Figure 4.3: Example trial for production task during training in experiment 4. Only the initial constituent was given at the beginning of each trial. The remaining constituents were shown underneath. By clicking on the words, participants could construct sentences. Participants could correct their productions before submitting their answers, by clicking the reset button. All provided words had to be used.

criterion was based on participants' performance during the reading task. If participants' accuracy score was lower than 66%, their data was excluded from further analysis. The more lenient exclusion criteria was motivated by the higher difficulty associated with this task compared to previous experiments: Participants need to remember unfamiliar words while also interacting with an unfamiliar word order. 15 reading trials were completed by participants.

The second training task involved production. Participants were provided with the clause-initial constituent along with three blank lines. The constituents to fill the blanks were provided as clickable buttons below the initial constituent and the three lines, as illustrated in Figure 4.3. All constituents had to be used — otherwise participants

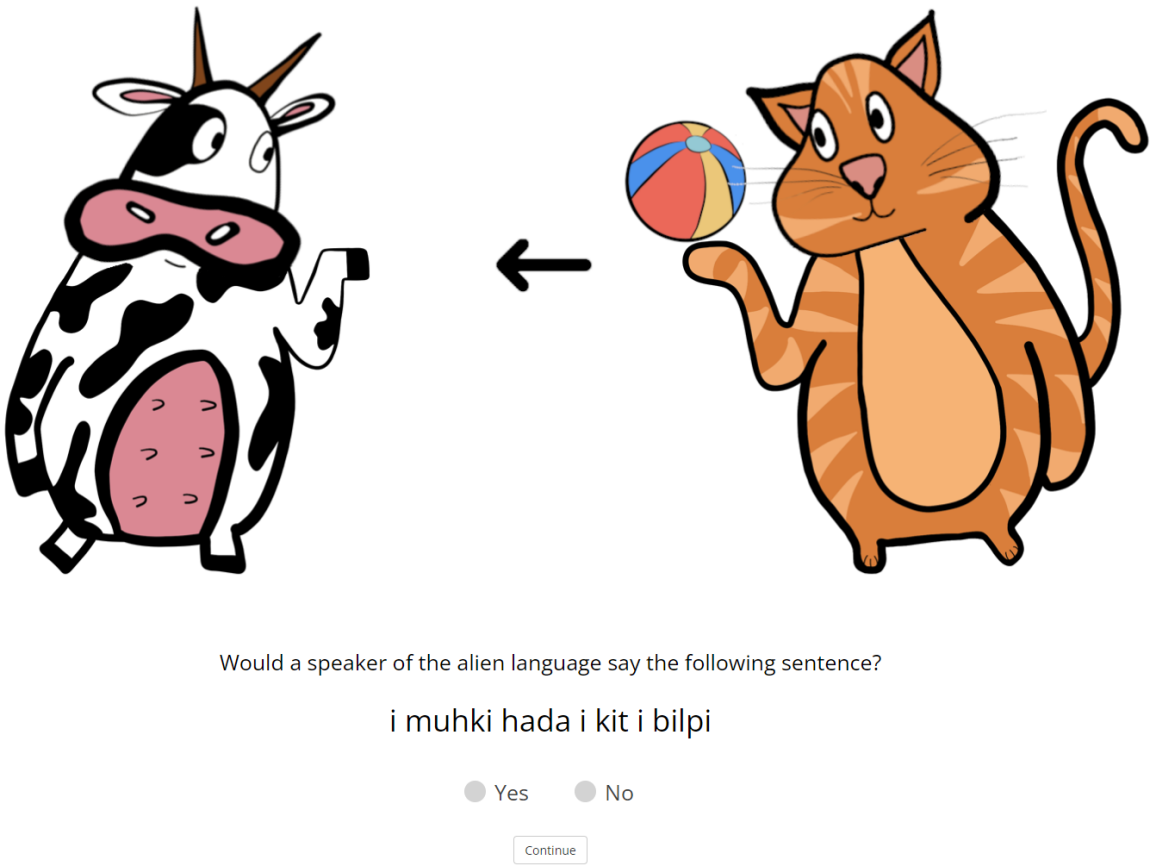


Figure 4.4: Example trial for judgement task during training in experiment 4. Participants had to determine whether the provided sentence could be produced by a speaker of the language.

could not proceed to the next trial. As in the reading trials, an image illustrating the sentence was provided above the sentence. The orientation of the image (i.e. flipped or non-flipped) was again determined by chance. The order of the unused constituents was randomised for each participant and trial. Participants could reset the constructed sentence before submitting their answer. Feedback was provided after each trial. The duration for which the feedback was shown on the screen was shorter (1500ms) if participants produced a correct sentence than if their production was erroneous (3000ms). Participants completed 30 trials of this type.

After participants completed the training, they were tested on their acquired knowledge of the language. The testing phase consisted of three parts. First, the words that were withheld from participants during training were introduced. The procedure was identical to the one in the training task (though without the testing component). The ditransitive verb *hada* '(to) give', however, was explained to participants by providing the English translation. This was necessary because the meaning could not be as easily

Part	Task type	*n* trials	Feedback
Lexical learning	Noun learning	12	n/a
	Noun testing	6	yes
	PP learning	6	n/a
	PP testing	3	yes
Syntactic learning	Reading	15	yes
	Production	30	yes
Syntactic testing	Production	23	no
	Judgement	24	no

Table 4.3: Summary of experimental procedure in experiment 4. The experiment consisted of three distinct phases. First, lexical items were taught to participants. Afterwards, the word order of the language was introduced to participants with the help of two tasks. Finally, participants' knowledge of the word order of the language was tested with again two tasks.

visually illustrated as the one of the nouns and PPs. The second part consisted of a production task akin to the one in training — the only exception was that no clause-initial constituent was provided and no feedback was given. Participants completed 23 trials of this type. The final task was a judgement task where participants had to judge whether a speaker of the alien language would use the shown sentence by clicking either 'Yes' or 'No' (Figure 4.4). A total of 24 judgement trials were included in the experiment. On a par with all previous tasks, an image was provided in each production and judgement trial that depicted the scene described by the stimulus sentences. The orientation of the image was randomly determined. Again, no feedback was provided. At the end of the experiment, participants were asked to fill in a questionnaire which gathered information on the strategies they employed to produce and judge sentences in the alien language. Furthermore, the questionnaire asked whether participants had noticed the case markings *-ki* and *-pi* and what language background participants have. The experimental procedure is summarised in Table 4.3.

4.3.2 Predictions

As in the previous experiments, the hypothesis investigated in this experiment is that higher variability of different clause-initial constituents — subject, direct object, and adjunct — will lead to better learning of V2. On a par with experiments 1 & 2 reported in Chapter 2, learning of a V2 grammar was operationalised as generalising the verb placement rule and the flexibility of the clause-initial constituent.[3] In other words, learners need to extrapolate (i) that the verb is realised in the second position irrespective of the

[3]The verb placement was not included in the operationalisation of V2 in the pre-registration of experiment 3. This is problematic: the realisation in the second position constitutes one of the defining features of V2. I therefore included verb placement as part of an exploratory analysis.

clause-initial constituent and (ii) that different constituent types — with the exception of the finite verb — can occupy the clause-initial position. In conjunction with the hypothesis under scrutiny, three predictions were derived from this operationalisation. First, participants should produce sentences with verbs in second position above chance. No differences between conditions are predicted to arise as verbs were always placed in second position during training in all conditions. This prediction therefore functions as sanity check. Second, participants in the uniform condition should produce or accept V2 sentences with initial novel constituents at a higher rate than participants in the skewed conditions. Finally, participants in the uniform condition should be better as discriminating grammatical V2 sentences with novel types of initial constituents from ungrammatical V3 sentences. Participants have not been familiarised with either of these sentence types. If participants accept the former but not the latter, I have evidence that participants generalise V2. In addition to this same basic hypothesis, given the findings of experiment 3, I will also be on the look-out for differences between the skewed conditions. In particular, the adjunct-initial skewed condition may again be learnt better. If this is the case, I will discuss the implications for my hypothesis.

4.3.3 Results

The data obtained in experiment 4 was analysed in accordance with the procedure used for experiments 1–3. That is, the R (R Core Team 2022) packages ggplot2 (Wickham 2016) and lme4 (Bates et al. 2015) were used for plotting and statistical analysis, respectively. Besides, Wald tests were used for obtaining *p*-values of model coefficients. The standard alpha level of 0.05 was used to determine significance.

4.3.3.1 Hypothesis-confirming analysis

Figure 4.5 depicts the proportion of V2 sentences produced by participants in each of the four conditions. According to the first prediction, participants should produce verbs in second position at a higher than chance level. Furthermore, no differences between the four conditions are expected. I fitted a mixed-effects logistic regression model to all produced sentences. Verb position was used as dependent variable (V2 = 1, non-V2 = 0). The model included CONDITION as fixed effect and by-participant and by-item random intercepts. CONDITION was treatment coded with the subject-dominant condition as the baseline. The model revealed that participants in the subject-dominant condition produced V2 sentences more often than expected by chance ($\beta = 7.38$, SE = .79, $p = 3.97 \times 10^{-27}$). To determine whether the other conditions also placed verbs in second position at a higher rate than expected by chance, three additional models with identical effect structure were fitted to the same data with adjunct-dominant, object-dominant and uniform, respectively, as the baseline. The models indicated that verbs were placed in second position at higher than chance level in all three conditions (object-dominant: $\beta = 6.33$, SE = .76, $p = 8.37 \times 10^{-23}$; adjunct-dominant: $\beta = 4.74$, SE = .74, $p = 3.30 \times 10^{-15}$; uniform: $\beta = 4.95$, SE = .70, $p = 4.19 \times ^{-18}$). To compare the different conditions with respect to the rate of V2 productions with each other, the simple effects of condition were examined in the models. No difference was found for the

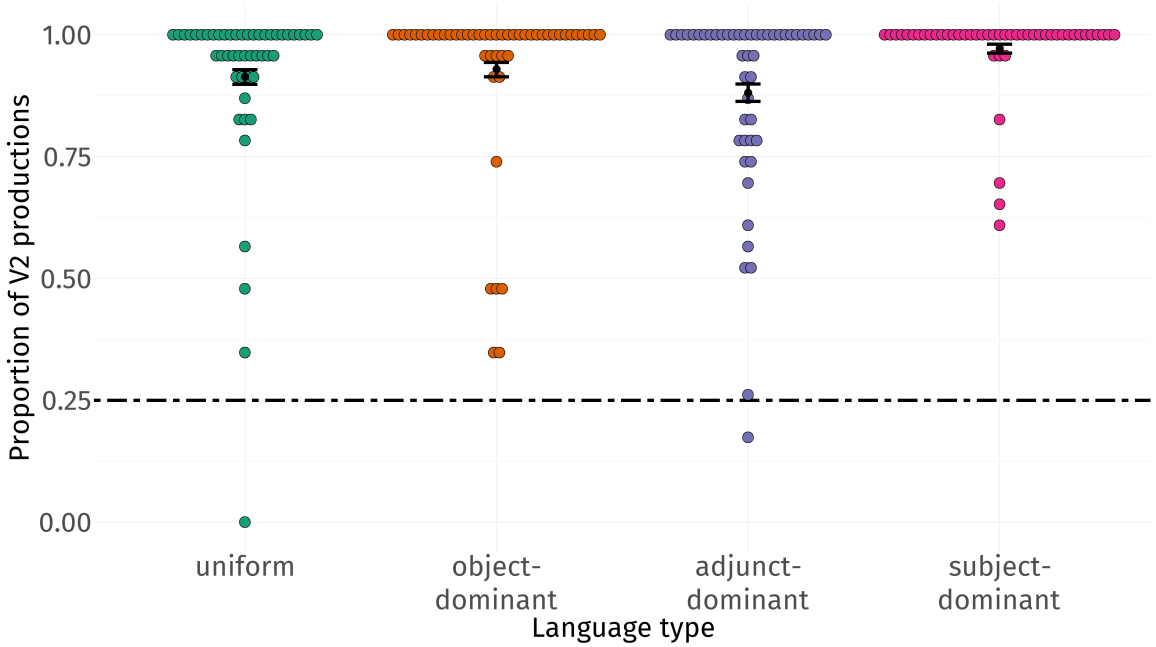


Figure 4.5: Proportion of produced V2 sentences by condition in experiment 4. Coloured dots represent the proportion of individual participants, black dots the means by condition. Error bars indicate bootstrapped 95% confidence intervals of the mean. The dotted line represents chance level. All four conditions produced V2 sentences at a high rate, greater than chance level, as predicted. However, the rate with which V2 sentences were produced differs across some conditions.

contrast between the subject-dominant condition and the object-dominant condition ($\beta = -1.06$, SE $= .85$, $p = .21$). Participants in the subject-dominant condition produced significantly more V2 sentences than participants in the adjunct-dominant condition ($\beta = -2.64$, SE $= .86$, $p = .002$) and uniform condition ($\beta = -2.43$, SE $= .86$, $p = .005$). The object-dominant condition, in turn, did not differ significantly from the adjunct-dominant condition ($\beta = -1.59$, SE $= .82$, $p = .053$) and the uniform condition ($\beta = -1.38$, SE $= .82$, $p = .09$). Also no significant differences in the rate of produced V2 sentences was indicated by the model for the comparison between the adjunct-dominant condition with the uniform condition ($\beta = .21$, SE $= .83$, $p = .80$). To summarise, while this analysis has revealed some unexpected differences between the conditions, the proportion of V2 sentences was significantly above chance, and high throughout all conditions.

The second prediction for the production data concerned the nature of the clause-initial constituent. I predicted that participants in the uniform condition should place more novel constituent types (i.e. indirect objects) in clause-initial position in V2 sentences than participants in the skewed conditions. I therefore examined only productions

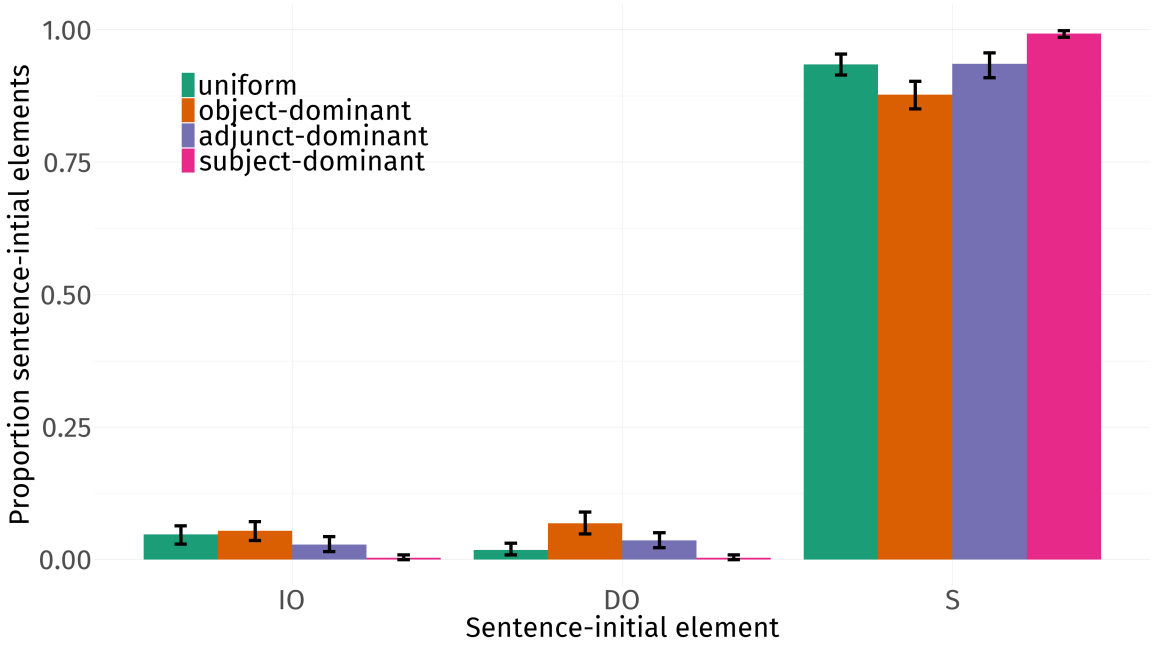


Figure 4.6: Proportion of familiar and novel constituents types in produced V2 sentences in experiment 4. Error bars indicate bootstrapped 95% confidence intervals. Participants produced predominantly familiar constituent types (i.e. subjects) in clause-initial position. The analysis showed no significant differences between conditions, the prediction was therefore not confirmed.

that involved indirect objects. Figure 4.6 shows the proportion of different types of clause-initial constituents (indirect objects, direct objects, and subjects) in sentences with V2 order produced by learners. While subjects and direct objects constituted familiar constituent types, indirect objects were a novel type of constituent for participants. Note that Figure 4.6 includes only cases where indirect objects were lexically novel (i.e. the lexical item was not seen during training). This is a strict test of generalisation since lexically familiar indirect objects could be produced in initial position simply because participants have previously seen them in that position. No comparable restrictions were applied to the lexical novelty of both subjects or objects; an insufficient number of test items would have remained for the analysis. Participants in all four conditions displayed a very strong preference for placing subjects in clause-initial position. However, to assess whether production patterns for indirect objects differed across conditions, a mixed-effects logistic regression model was fitted to all V2 sentences that included a lexically novel indirect object. Each sentence was coded for whether an indirect object ($= 1$) or some other constituent ($= 0$) was placed in clause-initial position. The model included CONDITION as a fixed effect and by-participant and by-item random intercepts. CONDITION was treatment coded with the subject-dominant condition as baseline. The model revealed that neither participants in the object-dominant-condition ($\beta = 1.70$,

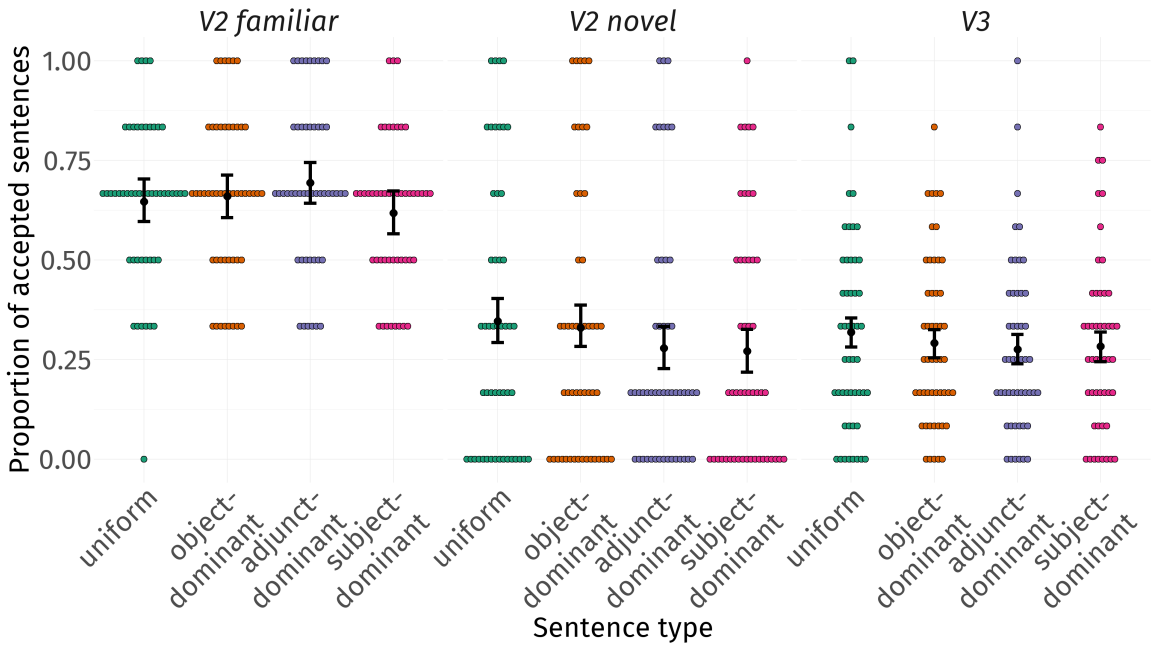


Figure 4.7: Acceptance rates of V2-familiar, V2-novel and V3 sentences by condition in experiment 4. V2-familiar encompasses all sentences with clause-initial subjects, objects and adjuncts. The category V2-novel comprises sentences with clause-initial indirect objects. V3 sentences include all sentences with V3 order. Coloured dots represent participants' mean acceptance rates, black dots the mean of the means. Error bars represent 95% confidence intervals. None of the two predictions for the judgement data were confirmed.

SE = 2.16, p = .43), in the adjunct-dominant condition (β = 1.35, SE = 2.24, p = .55) nor in the uniform condition (β = 1.97, SE = 2.14, p = .36) placed significantly more novel constituents in clause-initial position than those in the subject-dominant condition. To directly compare the other conditions, further models with the same effect structure were fitted to the same data but with different baselines of CONDITION. The models indicated no differences across the conditions in placing novel indirect objects in initial position (object-dominant condition as baseline: adjunct-dominant: β = −0.35, SE = 1.51, p = .81; uniform: β = .28, SE = 1.37, p = .84; adjunct-dominant as baseline: uniform: β = .63, SE = 1.48, p = .67). Hence, the second prediction for the productions was not borne out in the data.

Turning to the judgement data, as in previous experiments, a new factor SENTENCE TYPE was created. Recall that in this case though, I have both novel and familiar lexical items in each category. V2 sentences with clause-initial subjects, direct objects or adjuncts that are either lexically familiar or lexically novel were coded as *V2-familiar*. V2 sentences with clause-initial indirect objects — lexically familiar or lexically novel — were coded as *V2-novel*. As before, all V3 sentences were coded as *V3*, independent of

the clause-initial constituent or its lexical novelty or familiarity. As pre-registered, lexically novel elements were not treated differently in the analysis of the judgement data, because I hypothesised that lexical novelty should affect production more than judgement. The ratings for each sentence type are illustrated in Figure 4.7. Two predictions were tested for the judgement data: First, participants in the uniform condition should be more likely to accept V2-novel sentences compared to participants in the skewed conditions. Second, participants in the uniform condition should be less likely to accept V3 sentences than participants in the skewed conditions. To evaluate these predictions, I fitted a mixed-effects logistic regression model to the V2-novel and V3 data. The model included CONDITION and SENTENCE TYPE as fixed effects and an interaction term for both fixed effects. Furthermore, the model was fitted with by-participant and by-item random intercepts as well as by-participant random slopes for SENTENCE TYPE. The fixed effects were treatment coded with the subject-dominant condition and V2-novel sentences as reference levels. To assess the first prediction, I examined the simple effect of CONDITION. The model indicated that neither the object-dominant condition ($\beta = .50$, SE $= .47$, $p = .28$), the adjunct-dominant condition ($\beta = .19$, SE $= .47$, $p = .69$) nor the uniform condition ($\beta = .62$, SE $= .47$, $p = .18$) were more likely to accept V2-novel sentences than the subject-dominant condition. By refitting the model with different baselines of CONDITION, while keeping the effect structure otherwise constant, the other conditions were directly compared. Again, there were no differences between the conditions in the likelihood of accepting V2-novel sentences (object-dominant vs. adjunct-dominant: $\beta = -0.32$, SE $= .46$, $p = .49$; object-dominant vs. uniform: $\beta = .12$, SE $= .46$, $p = .80$; adjunct-dominant vs. uniform: $\beta = .43$, SE $= .46$, $p = .35$). The results thus show that the first prediction for the judgement data was not confirmed.

The second prediction for the judgement data — i.e. participants' discrimination between grammatical V2-novel sentences and ungrammatical V3 sentences should be better in the uniform condition — was assessed by investigating the interaction between CONDITION and SENTENCE TYPE. As already suggested by Figure 4.7, V3 sentences were as likely to be accepted by learners in the subject-dominant condition as V2-novel sentences ($\beta = .61$, SE $= .36$, $p = .09$). The interaction between V3 and the object-dominant condition was not significant ($\beta = -0.43$, SE $= .48$, $p = .37$), nor was the interaction between V3 and the adjunct-dominant condition ($\beta = -0.20$, SE $= .49$, $p = .69$) nor was the interaction between V3 and the uniform condition ($\beta = -0.42$, SE $= .48$, $p = .39$). When learners in the object-dominant condition were directly compared to those in the adjunct-dominant condition and uniform condition, no difference was found (adjunct-dominant: $\beta = .23$, SE $= .48$, $p = .63$; uniform: $\beta = .01$, SE $= .47$, $p = .98$). Similarly for the direct comparison of the adjunct-dominant condition and uniform condition, no difference in the discrimination between V2-novel and V3 sentences was found ($\beta = -0.22$, SE $= .48$, $p = .64$). In summary, the second prediction for the judgement data was also not confirmed.

4.3.3.2 Exploratory analysis

When the distribution of novel constituent types in participants' productions was analysed in §4.3.3.1, only V2 sentences with lexically novel indirect objects were

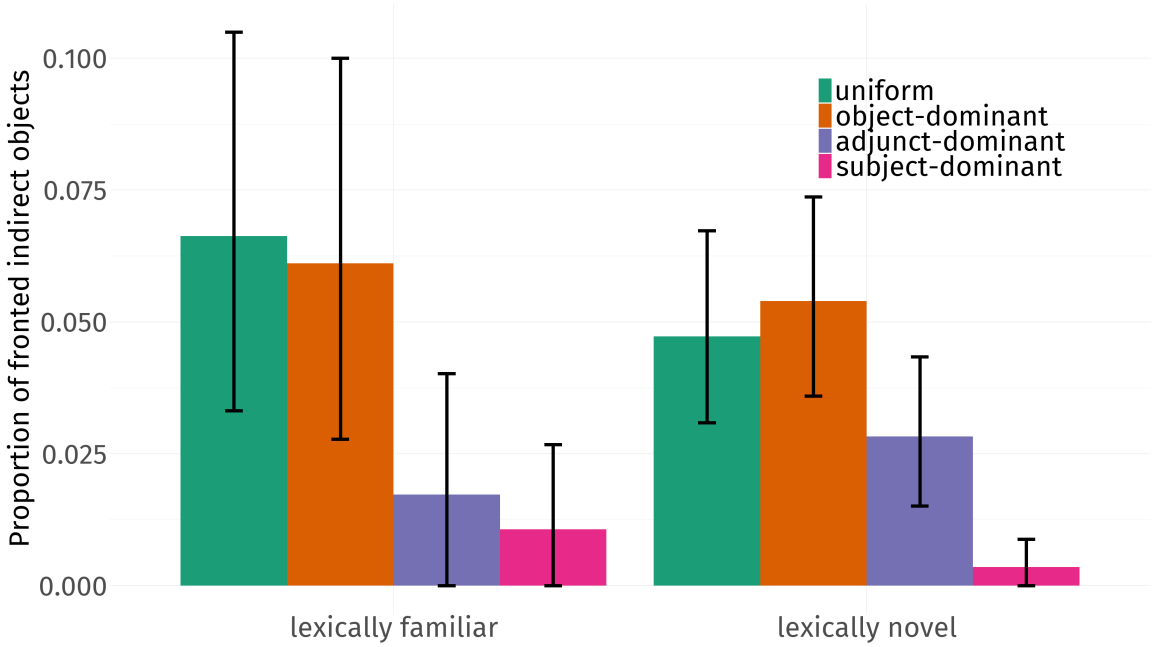


Figure 4.8: Proportion of lexically familiar and lexically novel indirect objects produced in clause-initial position in V2 sentences by participants in experiment 4. Error bars indicate bootstrapped 95% confidence intervals.

taken into consideration. This restriction was introduced as a strict measure of the generalisation of XP-fronting. Figure 4.8 shows the proportion of lexically familiar and lexically novel indirect objects placed in initial position (note that the y-axis range is restricted to better illustrate any differences). To examine the rates with which lexically familiar indirect objects are placed in initial position and to determine whether these rates differ from the rates for initial lexically novel indirect objects, a mixed-effects logistic regression model was fitted to all V2 sentences. Sentences were coded for whether an indirect object (= 1) or another constituent occupied the clause-initial position. The model comprised CONDITION and LEXICAL NOVELTY as fixed effects as well as an interaction term for both fixed effects. The model further included by-participant and by-item random intercepts and by-participant random slopes for LEXICAL NOVELTY. Both fixed effects were treatment coded with subject-dominant and lexically-familiar as reference level. I first examined the simple effect of CONDITION to determine whether the conditions differed in their production of V2 sentences with initial lexically-familiar indirect objects. The model indicated that none of the conditions produced significantly more V2 sentences with clause-initial lexically-familiar indirect objects than those in the subject-dominant condition (object-dominant: $\beta = 1.43$, SE $= 2.18$, $p = .51$; adjunct-dominant: $\beta = .01$, SE $= 2.53$, $p = .99$; uniform: $\beta = 1.25$, SE $= 2.19$, $p = .57$). To assess whether differences exist between the amount of produced clause-initial lexically familiar and lexically novel indirect objects in the different conditions, the

interaction between CONDITION and LEXICAL NOVELTY was studied. Participants in the subject-dominant condition were equally likely to produce V2 sentences with clause-initial lexically novel indirect objects and V2 sentences with clause-initial lexically familiar indirect objects ($\beta = -0.91$, SE $= 2.78$, $p = .74$). Neither participants in the object-dominant condition ($\beta = .21$, SE $= 2.77$, $p = .94$), in the adjunct-dominant condition ($\beta = 1.46$, SE $= 3.07$, $p = .63$) or the uniform condition ($\beta = .77$, SE $= 2.75$, $p = .79$) were more likely to produce a sentence with clause-initial lexically familiar indirect object, as the non-significant interactions reveal. That is, participants did not display a different behaviour when it comes to the production of lexically familiar and lexically novel clause-initial indirect objects.

To summarise, while participants appear to have acquired the position of the verb in all conditions, no other predictions with respect to the generalisation of the V2 property were confirmed in §4.3.3.1. The failure to confirm any other predictions raises the question of whether participants were able to learn any other properties of the language they were trained on. To scrutinise this question more closely, I conducted another exploratory analysis investigating the acceptance rates for V2-familiar sentences. If participants did not learn the language, this should become visible in the judgements of V2-familiar sentences. This sentence type featured exclusively familiar constituent types in clause-initial position. Figure 4.7 above illustrates the acceptance rate for V2-familiar sentences in all four conditions. This plot suggests that the acceptance rates were equally high for V2-familiar sentences in all conditions. This was confirmed in a mixed-effects logistic regression model fitted to V2-familiar sentences. The model included CONDITION as fixed effect and by-participant and by-item random intercepts. CONDITION was treatment coded with the subject-dominant condition as baseline. The model did not indicate a significant difference between the subject-dominant condition and the object-dominant condition ($\beta = .17$, SE $= .19$, $p = .38$), the adjunct-dominant condition ($\beta = .34$, SE $= .19$, $p = .08$) and the uniform condition ($\beta = .09$, SE $= .19$, $p = .62$).

Although these results suggest that participants in all four conditions learnt the input language equally well, some doubts remain. The frequencies of different clause-initial constituent types depicted in Figure 4.6 revealed a heavily skewed distribution. Participants displayed a very strong preference for placing subjects in clause-initial position. In the light of this pattern, one might ask whether the high acceptance rates of V2-familiar sentences were driven by high ratings for subject-initial V2 sentences. Figure 4.9 breaks down the acceptance rate for all four constituent types that occurred in clause-initial position in the judgement task. As can readily be gathered from the plot, the acceptance rates differ quite drastically for different clause-initial constituents and word orders. Subject-initial V2 sentences received the highest ratings in all four conditions. In fact, participants in the subject-dominant condition performed at ceiling: all subject-initial V2 sentences were accepted without exception. By contrast, the ratings for adjunct-initial V2 sentences are generally lower than those for subjects. Ratings for V2 sentences with initial direct and indirect objects exhibited the lowest acceptance rates. If participants did indeed learn the language they were trained on, they should be able to discriminate between grammatical V2 and ungrammatical V3 sentences with both

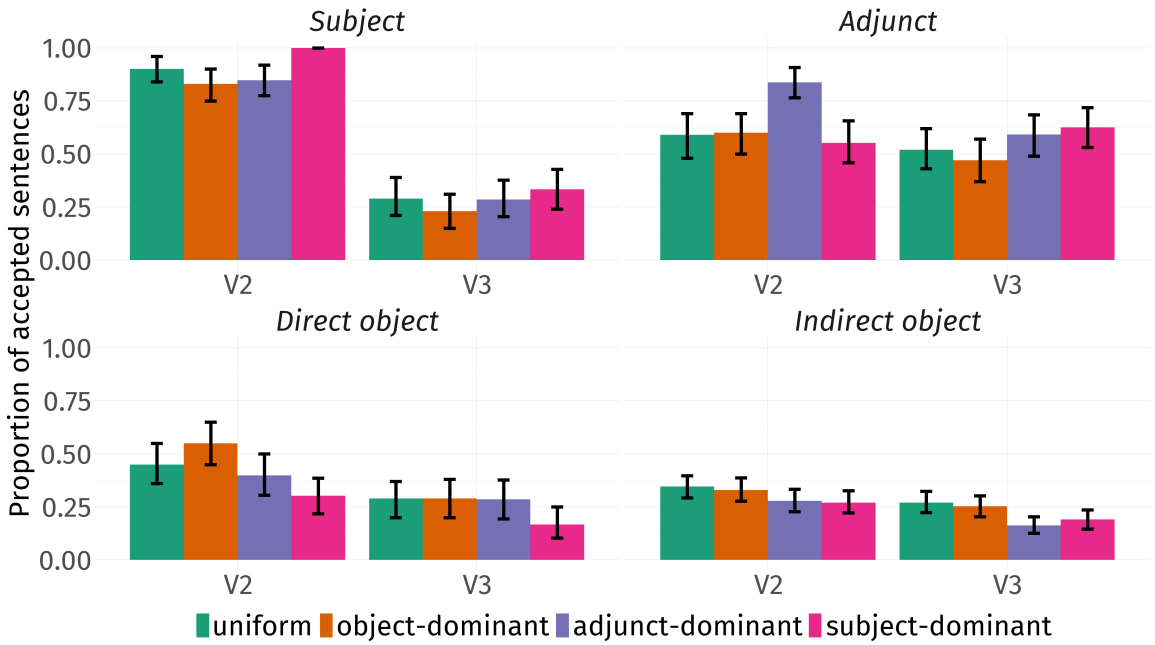


Figure 4.9: Acceptance rates of V2 and V3 sentences with different clause-initial constituents by condition in the judgement task in experiment 4. Error bars indicate bootstrapped 95% confidence intervals. Subjects, direct objects and adjuncts were familiar from training. Participants were not exposed to indirect objects during training, they thus constitute a novel constituent type. The ratings for indirect objects are included for reference only.

subjects *and* non-subjects in initial position.[4] However, the difference between V2 and V3 sentences was not as pronounced for non-subject constituent types. I thus analysed judgements for sentences with initial adjuncts and direct objects (i.e. the familiar non-subject constituent types). Two mixed-effects logistic regression models were fitted to these two sentence types respectively. The models included CONDITION and WORD ORDER as fixed effect as well as an interaction term for both. The model included by-participant and by-item random intercepts and by-participant random slopes for WORD ORDER. Both fixed effects were treatment coded with V2 and subject-dominant as baseline. Focusing first on adjunct-initial sentences, the intercept of the model suggests that participants in the subject-dominant condition rated adjunct-initial V2 sentences at chance level ($\beta = .24$, SE $= .24$, $p = .33$). The model also indicated that participants in adjunct-dominant condition ($\beta = 1.57$, SE $= .39$, $p = 5.53 \times 10^{-5}$) were more likely to accept adjunct-initial V2 sentences than participants in the subject-dominant condition. No differences were observed between participants in the subject-dominant condition

[4]This is reminiscent of my operationalisation of learning a V2 grammar. Crucially however, learning the input language does not equate to learning a V2 grammar. Participants do not extrapolate the verb placement and XP-fronting to novel contexts in this case.

and the object-dominant condition ($\beta = .22$, SE $= .34$, $p = .51$) and the uniform condition ($\beta = .17$, SE $= .34$, $p = .60$). Participants in the subject-dominant condition were equally likely to accept adjunct-initial V3 and V2 sentences ($\beta = .34$, SE $= .31$, $p = .28$). The non-significant interaction between V3 and the uniform condition ($\beta = -0.66$, SE $= .44$, $p = .13$) further suggests that participants in the uniform condition were also not able to discriminate between adjunct-initial V2 and V3 sentences. This contrasts with participants in the object-dominant condition and adjunct-dominant condition who were both significantly more likely to discriminate between adjunct-initial V2 and V3 sentences, as indicated by the interaction terms (object-initial: $\beta = -0.93$, SE $= .44$, $p = .03$; adjunct-initial: $\beta = -1.72$, SE $= .48$, $p = 2.92\times10^{-4}$). Taken together, these findings indicate that only participants in the adjunct-dominant condition and the object-dominant condition were able to correctly discriminate between grammatical and ungrammatical adjunct-initial sentences.

Next, I examined the model for object-initial sentences. Participants in the subject-dominant condition rated object-initial V2 sentences significantly below chance ($\beta = -1.24$, SE $= .36$, $p = 5.99\times10^{-4}$). The model showed that only participants in the object-dominant condition were more likely to accept object-initial sentences than participants in the subject-dominant condition ($\beta = 1.51$, SE $= .49$, $p = .002$). All remaining conditions were equally likely to accept object-initial V2 sentences (adjunct-dominant: $\beta = .62$ SE $= .49$, $p = .20$; uniform: $\beta = .94$, SE $= .48$, $p = .052$). Participants in the subject-dominant condition were less likely to accept object-initial V3 sentences compared to V2 sentences ($\beta = -1.05$, SE $= .42$, $p = .01$). The non-significant interactions between V3 and all other conditions indicates that participants in these conditions were not able to discriminate between object-initial V2 and V3 sentences (object-dominant: $\beta = -0.49$, SE $= .55$, $p = .37$; adjunct-dominant: $\beta = .33$, SE $= .55$, $p = .55$; uniform: $\beta = 0.05$, SE $= .54$, $p = .92$). These results suggest that participants in the subject-dominant condition were the only ones to discriminate between grammatical and ungrammatical object-initial sentences.

4.3.4 Discussion

Experiment 4 set out to replicate experiment 3 with a fully-artificial language. In experiment 3, I found that the adjunct-dominant condition, rather than the uniform condition, produced the best learning (and generalisation) results. Due to the combination of English vocabulary with an artificial V2 grammar in experiment 3, two possible explanations were suggested for the observed learning advantage of the adjunct-dominant condition: Participants' native language (i.e. English) might have interfered — positively in the case of the adjunct-dominant condition and negatively in the case of the object-dominant condition. Alternatively, a large proportion of adjuncts in clause-initial position might actually foster the acquisition of a V2 grammar. To explore the role of L1 influence, experiment 3 was repeated with a fully-artificial language. The hypothesis investigated here remained unaltered compared to experiment 3. High variability in the grammatical functions attested in the clause-initial position was hypothesised to foster the acquisition of a V2 grammar. This was operationalised as better extrapolation of

the verb placement rule and XP-fronting to a novel context. Four different distributions were compared to evaluate the hypothesis: a uniform distribution and three skewed distributions, a subject-dominant distribution, an object-dominant distribution and an adjunct-dominant distribution. After participants were trained on the language, I measured extrapolation (or generalisation) of the verb position and the clause-initial position by asking participants to produce and judge sentences that featured a novel constituent type in initial position, namely indirect objects. Using only this one novel constituent type is more limited compared to experiment 3, but can be considered as stronger test than the complex adjuncts used in addition to indirect objects in experiment 3.

The analysis of the productions and judgements provided by participants does not suggest that the language was learnt in any of the conditions. Although participants were able to place verbs in the correct position in their productions, they showed a substantial subject bias when it comes to the clause-initial constituent. The exploratory analysis of the production data further indicated that this cannot be attributed to the nature of the lexical elements (i.e. whether they were familiar from training or not). In the cases where novel constituent types were placed in the clause-initial position, no difference between the conditions manifested, contrary to my predictions. A similar picture emerged from the judgement data. Participants in all conditions were equally likely to accept V2 sentences with novel initial constituent types. Moreover, participants failed to discriminate between grammatical V2 sentences with novel constituent types in initial position and ungrammatical V3 sentences. These results thus point towards the conclusion that participants did not learn the V2 grammar. In other words, participants failed to generalise that the clause-initial position does not exhibit a fixed association with (a) particular grammatical function(s). The exploratory analysis further showed that participants even failed to learn the language they were trained on. Although the analysis of V2 sentences with familiar constituent types in initial position seems to suggest that participants actually learnt the language, this conclusion does not hold up under scrutiny. When the acceptance rates for V2 and V3 sentences with individual clause-initial constituents are compared, it becomes evident that only subject-initial V2 and V3 sentences could be reliably distinguished. For object-initial and adjunct-initial sentences, only some conditions could discriminate between grammatical V2 and ungrammatical V3 sentences. This is particularly striking for the object-dominant condition: Even though the acceptance rate for object-initial V2 sentences was higher than in other conditions, participants in this condition still failed to discriminate between V2 and V3 sentences.

The failure of learners to acquire a V2 grammar is particularly unexpected in the light of Getz's (2018) study: Even though the extent to which extrapolation occurred was not measured by Getz (2018), there participants at least learnt the language they were trained on. Why then did participants not learn this V2 language? Variability might provide an explanation. Recall from §1.5.1 that variability does not constitute a homogeneous concept, rather multiple types can be distinguished. Two of those types were numerosity and heterogeneity. Numerosity denotes the number of training items learners encounter during learning. This might explain the differences between the experiment in Getz (2018) and this experiment: While participants underwent 126

training trials in Getz (2018), only 45 training trials were completed by participants in my experiment. Even participants in experiments 1–3 were required to complete 90 training trials. Additional training might have been required to acquire the V2 system. The fact that participants were able to learn the placement of the verb, as indicated by the production data, does not contradict this explanation. Verbs consistently appeared in second position in the training data. It is therefore to be expected that participants perform better at verb placement than at other aspects of the language. That they fronted and accepted fronted subjects, suggests they may have assumed the languages was generally SVO. In other words, the relatively small number of training trials may have been insufficient to over-ride a prior assumption of SVO in these participants. It is worth noting that determining the appropriate number of training trials in an experiment like this is tricky: too many training trials might result in ceiling performance, obscuring differences between participants. It seems clear that here, though, I likely did not include enough.

Perhaps more interestingly, the second type of variability, i.e. heterogeneity, could have played a role as well. The number of distinct lexical items used in the current experiment was low. All words within a class were taken from the same semantic space (e.g. animals). In addition, all PPs were formed with the same preposition. In experiment 3 by contrast, participants saw a much wider range of items, both between categories and within categories. The lack of variability might thus have affected learners' performance negatively.[5] Such an explanation is further supported by the findings of Poletiek & van Schijndel (2009) and Schiff et al. (2021) who showed that participants learning a finite-state grammar benefited from more heterogeneous input but not from additional training trials. Note though that Getz (2018) used a smaller lexicon than I did: three nouns, two verbs and two adverbs. In Getz (2018) however, verbs bore inflectional markers, which appeared to be beneficial for learners (cf. §2.2.2). Moreover, the verb position was alternating between V2 and V-final. As argued in §1.5.2, such alternations may provide additional evidence to learners for the acquisition of a V2 grammar. The fact that participants were able to learn the input language in Getz (2018) could thus be attributed to the presence of additional evidence in the input. Future work using the paradigm I have developed could include additional such cues (e.g. alternating verb positions).

Expanding the lexicon to increase heterogeneity is harder to achieve in a short experimental session. A substantial increase in the lexicon of the artificial language would require an extended training period, potentially over multiple days. Although artificial languages *can* be successfully taught over multiple days (Hudson Kam & Newport 2005, 2009, Thompson & Newport 2007), this is not the only option. Fortunately, Shapiro & Steinert-Threlkeld (2023) offer an intriguing alternative: iconic artificial language learning. In this new paradigm, lexical items of the language are replaced by icons, i.e. pictographic representations of the concepts denoted by the lexical items. Shapiro & Steinert-Threlkeld (2023) provide a proof-of-concept by replicating the results of previous studies investigating the word order in the noun phrase (Culbertson & Adger

[5]This would also align with the second explanation for the results in experiment 3, cf. §4.1.

2014, Martin et al. 2019, 2020). The novel experimental paradigm offers the advantage that participants do not need to learn the lexical items first before the syntactic structure can be learnt. This is concomitant with the absence of restrictions on the number of lexical items that can be used. Such diversity is particularly relevant when variability of lexical items is required — as it is the case in the present thesis. The design also offers the added advantage that different linguistic populations can be tested with the same language thereby alleviating problems of typologically unrepresentative samples of tested populations (Shapiro & Steinert-Threlkeld 2023: 146).[6]

In this section, I have presented the results of a fully-artificial language learning experiment. The goal of the experiment was to replicate the findings of experiment 3 (cf. §3.2). The analysis showed however, that learners failed to acquire the V2 language. In fact, participants even failed to learn the language they were trained on. In the following discussion, I argued that the lack of variability in the lexicon might have caused this. In experiment 5, I use a new paradigm, namely iconic artificial language learning, to train participants on a V2 grammar.

4.4 Experiment 5

The goal, hypothesis and predictions of experiment 5 remain identical to experiment 4. Other than the change to the nature of the language itself, I also dropped the subject-dominant condition. Participants in experiment 4 displayed a very strong initial subject bias suggesting that this condition is probably uninformative for my purposes.

4.4.1 Methods

The software used in experiment 5 (running the experiment and analysis) was identical the one in experiment 4. Since the hypothesis as well as the predictions are identical to those in experiment 4, I based my analysis on the preregistration for that study. The experiment received ethical approval from the ethics board of the Linguistics and English Language department of The University of Edinburgh (395-2223/1).

4.4.1.1 Participants

156 participants were recruited online after the experiment was advertised on Prolific. Participants received £4.34 as compensation for their participation. In-built Prolific prescreeners were used to restrict participation. The same geographical restrictions as in experiment 4 were set (i.e. USA, Canada, UK, Ireland, Australia & New Zealand). Similarly for the language background, participants had to have grown up monolingually with English as their earliest and only language to be eligible to participate. In addition, their primary language had to be English and their subjects at university could not include English language, English literature or languages more generally. As before, participants' approval rating for past participation on Prolific had to be at least 95%.

[6]Note that this advantage is not relevant for the current set of experiments. However, if non-English-speaking populations should be tested in the future, this advantage will become relevant.

Six participants were excluded from further analysis because they failed to meet the inclusion criterion based on their performance during training (cf. §4.4.1.3). One participant reported knowledge of Norwegian (a V2 language, see §1.2.2) and one participants admitted to have taken notes. However, both participants failed to meet the inclusion criterion and were therefore already excluded. 50 participants in each of the three conditions — i.e. the uniform condition, the object-dominant condition and the adjunct-dominant condition — were analysed.

4.4.1.2 Materials

Following Shapiro & Steinert-Threlkeld (2023), sentences for training and testing were constructed using an iconic artificial language. That is, the lexical items of the language were represented by icons instead of written words.[7] As aforementioned, the usage of icons instead of words has similar benefits to using a semi-artificial language: Participants do not need to learn the lexical items but can instead focus on learning the syntactic pattern of the language. The icons were picked such that the meaning be easily identifiable with participants' cultural background. Icons were divided into four semantically motivated classes: animate persons, inanimate objects, actions and environments. Animate persons denote professions or athletes (in a broad sense); inanimate objects were either foodstuff or objects of everyday life. Actions involved some degree of physical action, whereas environments described either the weather conditions or a location. These icon categories will henceforth be referred to as animate nouns, inanimate nouns, verbs and PPs, to highlight the similarities to the previous experiment. In the language that participants learnt, these four classes correspond to a specific syntactic function, namely subjects, direct objects, verbs and adjuncts. No movement marker was included for two reasons. First, sentence-medial adverbs such as *often* may not be easily identifiable when represented as icons. Second, movement markers in experiments 1–3 signalled to learners that the English variety they were learning differed from their native one. Given the absence of a comparably close relationship between participants' L1 and the language used in the present experiment, the inclusion of such indicators for movement was deemed less critical compared to previous experiments. The number of lexical items was not evenly distributed across classes: Whilst animate nouns, inanimate nouns and PPs contained 18 elements each, only 12 lexical items were included as verbs. The rationale for including a lower number of verbs was that the lexical variability in the clause-initial position was considered more important than the one in second position (i.e. verbs). Examples of icons used as animate nouns, inanimate nouns, verbs and PPs are provided in Table 4.4.

As in experiments 3 and 4, the distribution of clause-initial elements (subject, direct object and adjunct) was contingent on the experimental condition. All three constituent types were equally frequent in the clause-initial position in the uniform condition. In the object-dominant condition, direct objects were more likely to appear in the initial position. In the adjunct-dominant condition, adjuncts were the most frequent initial elements. The subject-dominant condition was dropped again from the design given

[7] All icons were taken from Flaticon.com.

Category	$Item_1$		$Item_2$	
Animate N		cyclist		diver
Inanimate N		bag		chocolate
Verb		(to) paint		(to) wipe
PP		in the mountains		on the beach

Table 4.4: Examples of icons used in experiment 5. Four categories of icons can be distinguished: animate nouns, inanimate nouns, verbs and PPs.

the overwhelming SVO-bias displayed by participants in that condition in experiment 4 (cf. §4.3.3).

I created 12 unordered sets of constituents used for training. 12 animate nouns, 12 inanimate nouns, 12 PPs and six verbs were selected from the whole lexicon. Each icon was included in one set only, apart from verbs which featured in two sets. The remaining icons in the lexicon were completely withheld from participants during training and were used to construct testing materials (see below). From each set, three different visual sentences were created, a subject-initial version, an object-initial version and an adjunct-initial version. A total of 36 different sentences were thus created by this method.

Three training blocks were devised for each condition. The training items for the uniform condition were created by assigning each of the sentences created from the 12 unordered sets to two of the three training blocks. Sentences were distributed evenly across blocks. No sentence was assigned to the same block twice. Moreover, the number of different clause-initial constituents was balanced within and across blocks. That is, subjects, direct objects and adjuncts accounted for equal proportions of initial constituent types (i.e. 33.3%) in the uniform condition. In the skewed conditions in contrast, 50% of all training sentences featured the dominant constituent type in initial position (i.e. direct objects or adjuncts). Note that this proportion is 10% lower than in experiments 3 and 4. This reduction was necessitated by the smaller number of training sets. Two different sentences from each training set were assigned to a training block. Due to the skew, a different method was used for the assignment of the sentences to the training blocks. The sentence with the dominating element in initial position was assigned to each of the three training blocks. One of the two sentences with non-dominant constituent types was assigned to two blocks, the other sentence to one block. The overall distribution of clause-initial constituents was maintained within and across training blocks in both skewed conditions. Table 4.5 summarises the number of different sentence types for each of the conditions. By Latin Squares, three training lists were created from the training blocks for each condition. The order of the training

Condition	Subject-initial	Object-initial	Adjunct-initial
Uniform	24	24	24
Object-dominant	18	36	18
Adjunct-dominant	18	18	36

Table 4.5: Distribution of subject-initial, object-initial and adjunct-initial sentences in each of the three conditions in experiment 5.

sentences in a block was randomised for each participant. The first block was used for a reading task and the two remaining blocks for a production task (cf. §4.4.1.3).

Two sets of testing materials were constructed — one for a production task and one for a judgement task. Materials for the former consisted of unordered sets of icons. The icons could be used by participants to form visual sentences (cf. §4.4.1.3 for more details). All of the sets contained a subject, a direct object and a verb. In half of the sets, the fourth constituent was an adjunct, whereas in the other half a novel constituent type was added. On a par with experiment 4, this novel constituent type was indirect objects. Eight sets with familiar constituent types only and eight sets with the novel constituent type were created. Half of the familiar sets contained subjects, direct objects and adjuncts that were not included in the training materials, they were thus lexically novel. The verbs were used in the training items, they were thus lexically familiar. The other half consisted of subjects, direct objects and adjuncts that were lexically familiar from the training. However, none of the icons previously co-occurred during training. The verbs in this type of set were lexically novel. In the icon sets with indirect objects, a similar procedure was followed. Half of the sets featured lexically unfamiliar subjects, direct objects and indirect objects, while in the other half, all were lexically familiar. Note that the verbs were lexically novel in all of those sets, as ditransitive verbs were not used during training. Two ditransitive verbs were included, *give* and *throw*. Apart from adjuncts, all lexically novel items were used at least once for the production materials. Only four of the six lexically novel adjuncts could be used as only four lexically-novel production sets were included.

The second set of testing materials was created for a sentence judgement task (cf. §4.4.1.3). Sentences were created by crossing three factors: VERB POSITION (V2 or V3), INITIAL CONSTITUENT (subject, direct object, adjunct, or indirect object) and LEXICAL ITEM (lexically familiar or lexically novel initial constituent). The number of testing sentences created for each factor combination depended on which levels were combined. For V2 and V3 sentences with initial subjects, direct objects and adjuncts, two lexically novel sentences and one lexically familiar sentence each were constructed. Combinations that involved indirect objects were more frequent in the materials. For combinations involving lexically familiar indirect objects in initial position, four sentences were created and five sentences for combinations that included lexically novel indirect objects. None of the icons previously co-occurred with any other icon except for verbs. Note that the lexical novelty of the remaining constituents was balanced across sentences.

Figure 4.10: Example trials for reading task during training in experiment 5. Constituents were revealed successively, i.e. *in the desert*, *eats*, *the diver* and *the chocolate*. Participants were then asked to click on one of the constituents, as instructed.

See Appendix C for a full list of all items. A total of 36 testing sentences were thus created.

4.4.1.3 Procedure

Participants could access the experiment through a web browser from their laptop or personal computer. On the first page of the experiment, participants were informed they would be learning a visual language. Following a general introduction, a comprehension question followed asking what participants would learn in the experiment. Five options where given, four foils and one correct answer. If participants selected the wrong answer more than once, they could not continue with the experiment and were asked to return their submission to Prolific without payment. The experiment proper consisted of two phases, a training phase and a testing phase. Participants were assigned randomly

to one of the three training lists per condition. Akin to my previous experiments, the training phase was divided into two parts, namely a sentence reading task and a sentence production task. Before participants began with the sentence reading task, they were informed that they would read sentences and would need to identify different constituents. More neutral instructions were adopted compared to experiment 3 and experiment 4 by changing the labels of the constituent types to the doer (i.e. subject), the thing affected by the action (i.e. object), the action (i.e. verb) and the environment (i.e. adjunct). To aid participants' understanding of the task, an English example sentence was provided for which each of the categories was identified. During each trial, the procedure was identical to experiment 4: an image depicting the scene described by the sentence was displayed. The image was randomly vertically flipped for each participant. This should prevent the induction of any biases reflecting the position of the actor and patient in the image and in the sentence. The sentence was revealed icon by icon, with a delay of 600ms. The time span between the revelation of icons was slightly increased compared to previous experiments to account for the more unusual character of the experiment. Once the sentence was fully visible, participants had to identify one of the constituents by clicking on the respective icon, as illustrated in Figure 4.10. Each of the constituent types was queried six times. Participants received feedback after each trial; if the choice was correct, the feedback was displayed for 1500ms, and if the wrong constituent was selected, feedback was shown for 4000ms. Again, the extended times should give participants sufficient time to process the unusual task. The performance on this task by participants was used as exclusion criteria. The same threshold as in experiment 4 was used. That is, participants had to select at least 66% of all constituents correctly to be included in further analysis (cf. §4.4.1.1). Participants completed 24 reading trials.

In the next part of the training, participants were asked to produce sentences in the language. The clause-initial constituent as well as three blank lines were provided on each trial, as Figure 4.11 exemplifies. Akin to experiments 2–4, participants' task was to complete the sentence by clicking on the icons that were shown in a triangle formation to the right of the empty lines. The order of the icons was randomised for each participant and all icons had to be used. Feedback was provided after each trial. If the order of the icons was correct, feedback was displayed for 1500ms. However, if the order was incorrect, feedback was shown for 4000ms. Incorrectly placed icons were highlighted by a red background. In addition, the correct order was shown underneath the icons. A total of 48 trials were completed by participants.

The second part of the experiment was the testing phase. Two different types of test were used. The first task was a production task, similar to the one participants had completed during training. However, no initial constituent was given and participants received no feedback after each trial. Participants had to complete 16 trials of this kind. The second part was again a judgement task. Participants had to judge whether a speaker of the language would use the given sentence. The meaning of each icon sequence was illustrated by an image. The orientation (i.e. whether the image was vertically mirrored or not) was randomised. Participants could chose between 'yes' or no' as in Figure 4.12. Again no feedback was provided to participants. After participants

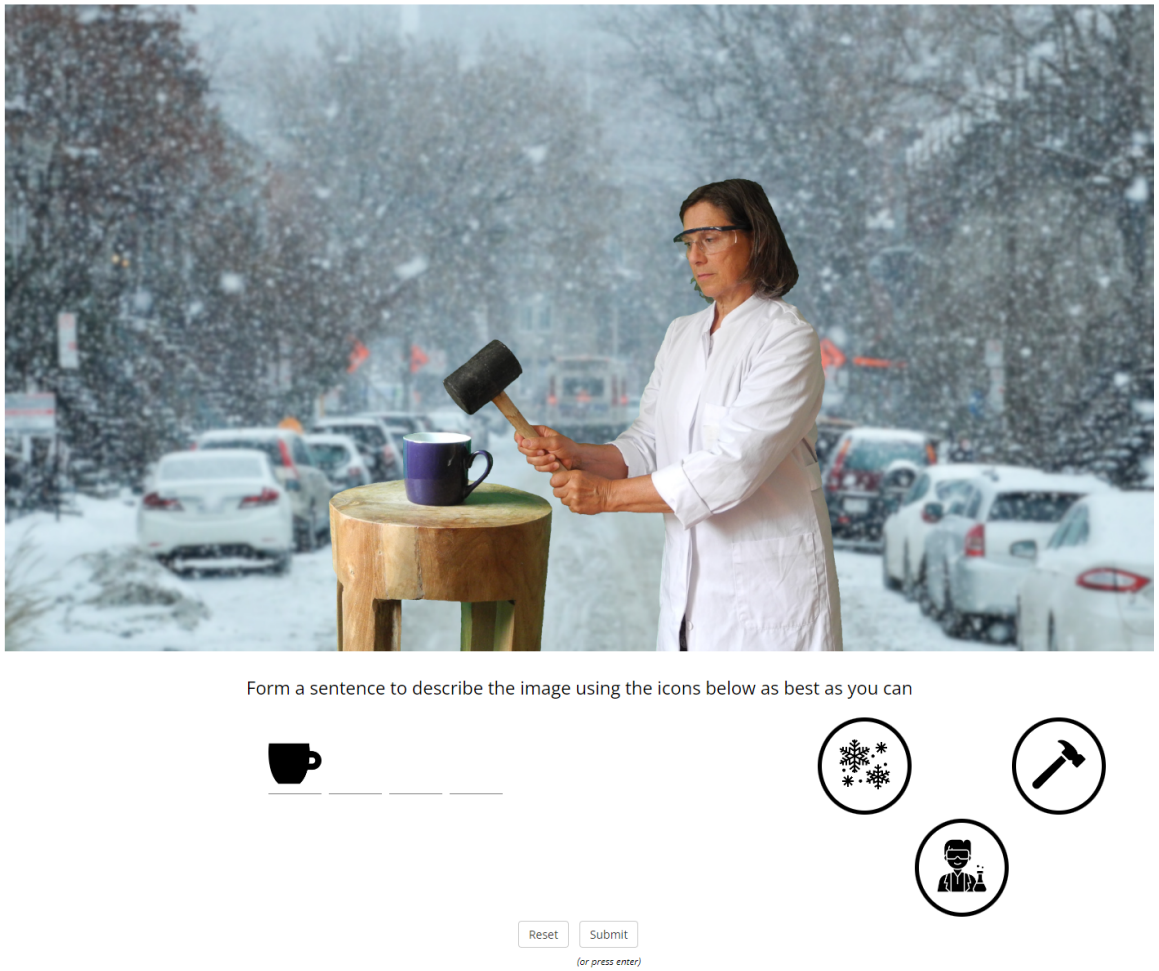


Figure 4.11: Example trial for production task during training in experiment 5. The clause-initial constituent was always provided to participants. The remaining constituents were given on the right of the empty line. By clicking on the icons, participants could insert them into the sentence. The order of the icons was randomised for each participant.

completed all 36 judgement trials, they were asked to translate three icon sequences into English. This task was included as test for whether participants correctly understood the icon sequences as sentences. Crucially, no image was provided to illustrate the meaning and none of the icons co-occurred in the same sentence before. The experiment concluded with a questionnaire, enquiring about participants' strategies for completing the experiment (including their strategies for judging ditransitive sentences) as well as their linguistic background. Table 4.6 summarises the procedure briefly.

Figure 4.12: Example trial for judgement task during testing in experiment 5. Participants had to decide whether the sentence illustrated by the image would be used by a speaker of the visual language.

4.4.2 Predictions

The hypothesis and predictions remained unaltered compared to experiment 4 (cf. §4.3.2). That is, higher variability in the clause-initial position should lead to better learning of V2. Combined with the previously used operationalisation of V2 — i.e. extrapolation of the verb placement rule and the flexibility of the initial constituent to novel structures — I derived the following predictions: First, verbs are placed in second position more often than expected by chance. No differences between conditions

Part	Task type	*n* trials	Feedback
Training	Reading	24	yes
	Production	48	yes
Testing	Production	16	no
	Judgement	36	no
	Translation	3	no

Table 4.6: Summary of experimental procedure in experiment 5. The experiment consisted of two distinct phases. First participants were trained on the word order of the language. Afterwards, they were tested on how well they have learnt the order. In addition, they had to translate sentences from the visual language into English.

are expected. Second, participants in the uniform condition should produce as well as accept sentences with novel constituent types in initial position at a higher rate than participants in the skewed conditions. Finally, participants in the uniform condition should be better at discriminating V2 sentences with novel constituent types in clause-initial position from ungrammatical V3 sentences.

4.4.3 Results

4.4.3.1 Hypothesis-confirming analysis

Figure 4.13 shows the proportion of verbs placed in the second position in participants' production by condition. I predicted that irrespective of the condition, participants should place verbs in second position above chance, with no difference between conditions. To verify this prediction, I fitted a mixed-effects logistic regression model to participants' production. All produced sentences were coded for whether the verb was realised in second position (=1) or not (=0). The model included condition as fixed effect as well as by-participant and by-item random intercepts. CONDITION (object-dominant, adjunct-dominant or uniform) was treatment coded with the object-dominant condition as baseline. The model showed that participants in the object-dominant condition placed verbs in second position significantly above chance ($\beta = 4.02$, SE = .51, $p = 1.32 \times 10^{-23}$). Learners in the adjunct-dominant condition did not differ from this ($\beta = -1.16$, SE = .62, $p = .06$), nor did learners in the uniform condition ($\beta = .02$, SE = .63, $p = .97$). The adjunct-dominant condition and the uniform condition were directly compared by fitting another model with identical effect structure to the same data with adjunct-dominant as baseline for CONDITION. The model again showed no significant difference ($\beta = 1.18$, SE = .61, $p = .053$). That is, the results from the production of V2 sentences match my predictions.

To assess the second prediction for the production data, I examined the constituent types in V2 sentences produced by participants. I predicted that participants in the uniform condition should place more novel constituent types in clause-initial position

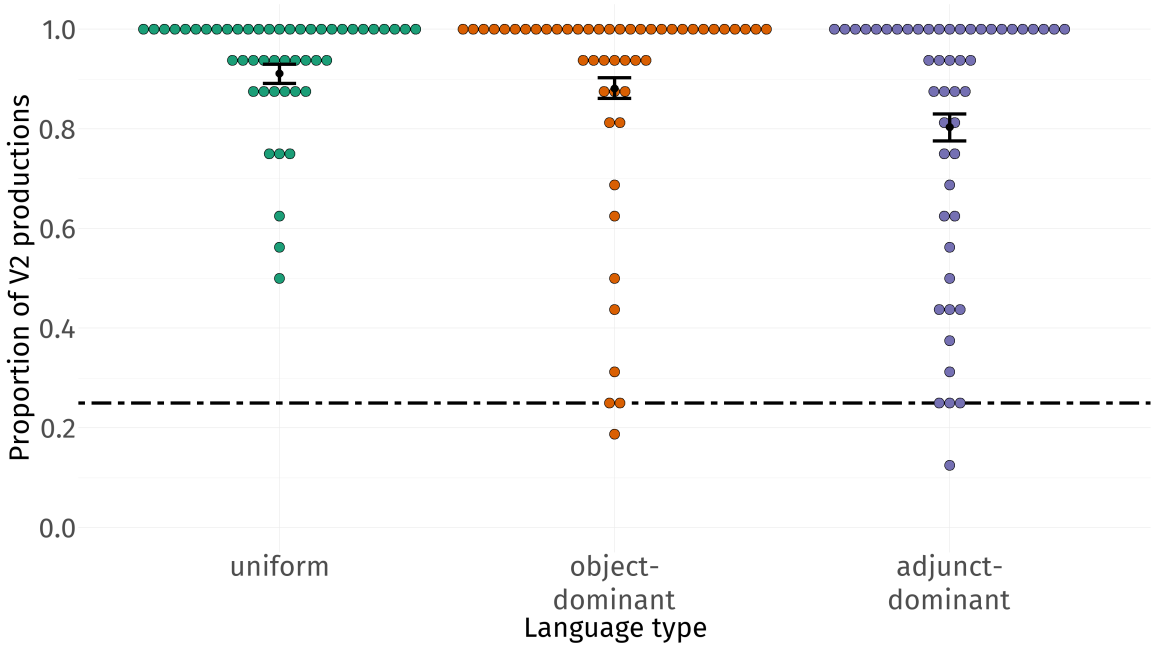


Figure 4.13: Proportion of produced V2 sentences by condition in experiment 5. Coloured dots represent the proportion of individual participants, black dots the means by condition. Error bars indicate bootstrapped 95% confidence intervals of the mean. The dotted line represents chance level. All three conditions produced V2 sentences at a high rate, greater than chance level. No differences could be observed between conditions, as predicted.

than those in the skewed conditions. Indirect objects constitute novel constituent types, while subjects, direct objects and adjuncts form the category of familiar constituent types (though adjuncts were not included in sets with indirect objects). Figure 4.14 visualises the proportion of different constituent types in clause-initial position in V2 sentences produced by participants. As in experiment 4 (cf. §4.3.3.1), only sentences formed from sets with lexically novel constituents — i.e. those that were unfamiliar from training — were included. The prediction was assessed with a mixed-effects logistic regression model. The dependent variable encoded whether a novel constituent type (=1) or a familiar constituent type (=0) was placed in initial position. The model included CONDITION as fixed effect. In addition, the model comprised by-participant random intercepts. By-item random intercepts were dropped due to singular fit. CONDITION was treatment coded with object-dominant as baseline. The model revealed that participants in the adjunct-dominant condition did not place more novel constituents in initial position than participants in the object-dominant condition ($\beta = .43$, SE $= 1.11$, $p = .70$). Learners in the uniform condition also did not produce significantly more V2 sentences with initial novel constituents than learners in the object-dominant condition ($\beta = -0.70$, SE $= 1.19$, $p = .56$). To directly compare the adjunct-dominant condition

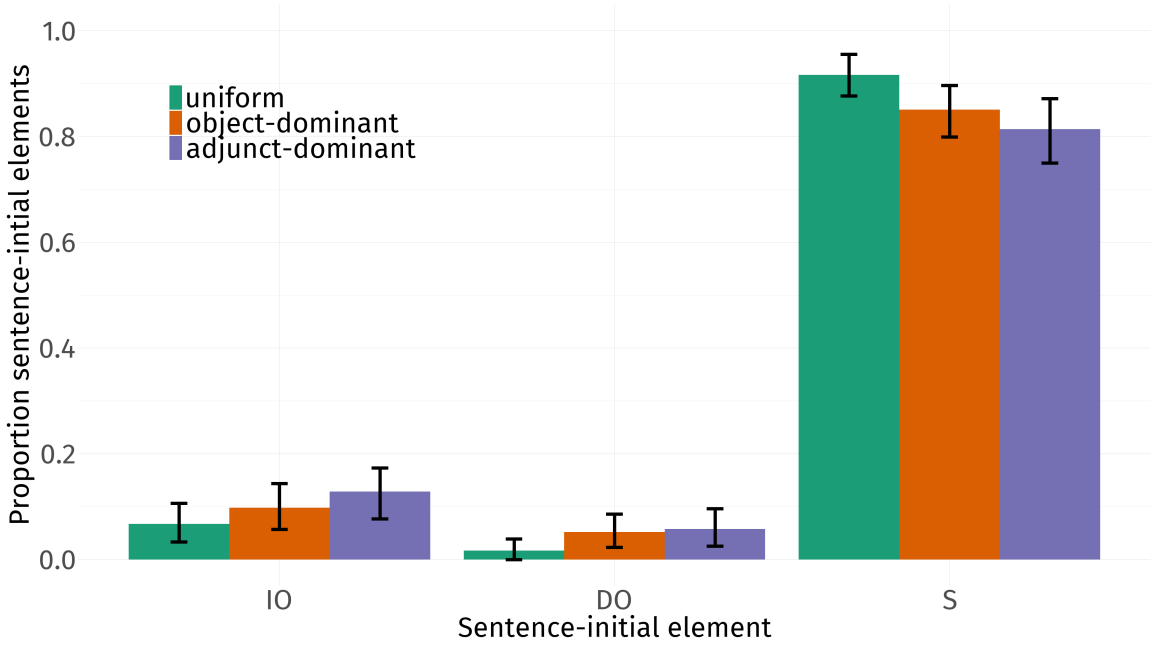


Figure 4.14: Proportion of familiar and novel constituent types in produced V2 sentences in experiment 5. Error bars indicate bootstrapped 95% confidence intervals. Similar to experiment 4, participants showed a preference for clause-initial subjects (i.e. a familiar constituent type). The analysis showed no significant differences between the three conditions in that novel constituents (i.e. indirect objects) were placed equally frequent in initial position. The prediction was therefore not confirmed.

with the uniform condition, another model with identical effect structure was fitted to the same data, the only difference being that the adjunct-dominant condition was used as baseline for CONDITION. The model results do not provide evidence for a significant difference between the two conditions ($\beta = -1.13$, SE $= 1.19$, $p = .34$). The findings do not match my predictions as no advantage of the uniform (or adjunct-dominant) condition was found.

I then examined the judgement data. A new factor SENTENCE TYPE was created by grouping together different levels of VERB POSITION and INITIAL CONSTITUENT. V2 sentences with initial subjects, direct objects and adjuncts were categorised as *V2-familiar*. V2 sentences with indirect objects in clause-initial position were subsumed under the label V2-NOVEL. All other sentences irrespective the initial constituent were grouped together as *V3*. The ratings for the different sentence types are shown in Figure 4.15. I formulated two predictions for the judgement data: participants in the uniform condition (i) should accept V2-novel sentences at a higher rate than participants in the skewed conditions and (ii) should be less likely to accept ungrammatical V3 sentences in contrast to the skewed conditions. In order to test the two predictions, a mixed-

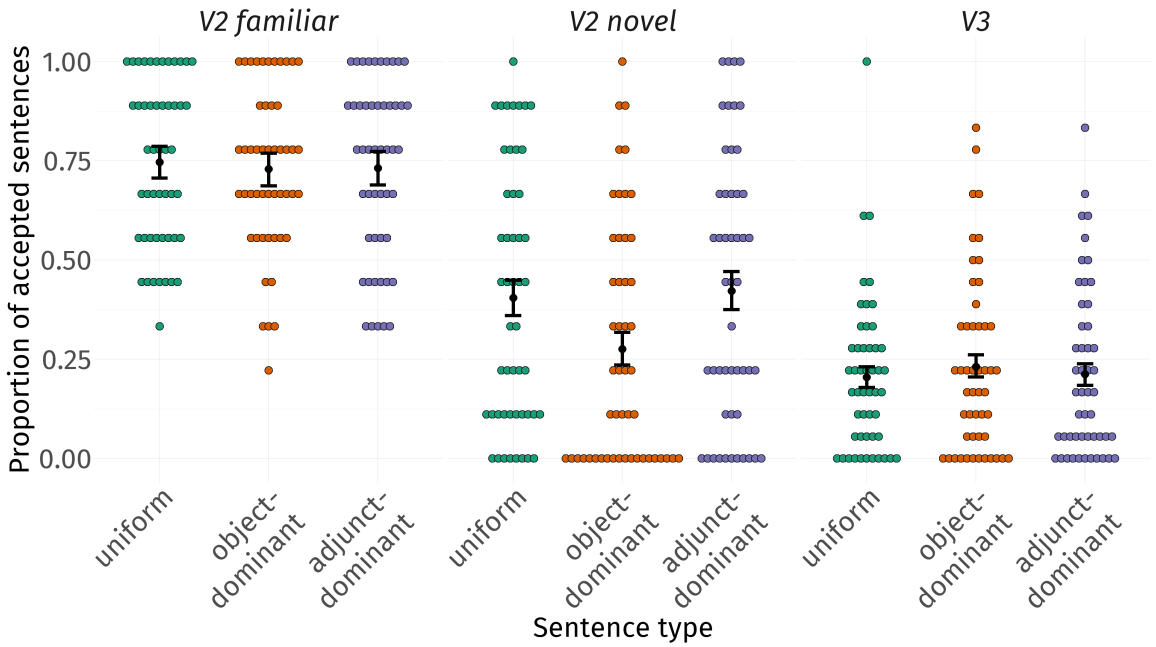


Figure 4.15: Acceptance rates for V2-familiar, V2-novel and V3 sentences by condition in experiment 5. V2-familiar sentences include V2 sentences with clause-initial subjects, direct objects and adjuncts. V2-novel sentences comprise V2 sentences with clause-initial indirect object, i.e. a constituent type that participants have not seen during training. V3 sentences subsume all V3 sentences irrespective of their clause-initial constituent. Coloured dots show the mean acceptance rate of individual participants, black dots the mean of the means. Error bars represent 95% confidence intervals. The predictions were confirmed for the contrast between the uniform condition and the object-dominant condition but not for the contrast between the uniform condition and the adjunct-dominant condition.

effects logistic regression model was fitted to the V2-novel and V3 data. The model included CONDITION and SENTENCE TYPE as fixed effects and an interaction term for both. Furthermore, the model comprised by-participant and by-item random intercepts as well as by-participant random slopes for SENTENCE TYPE. Both fixed effects were treatment coded with the object-dominant condition and V2-novel as baseline. The first prediction was assessed by examining the coefficients for CONDITION. Participants in the adjunct-dominant condition were more likely to accept V2-novel sentences than those in the object-dominant condition ($\beta = 1.16$, SE $= .46$, $p = .01$).[8] The model revealed further that participants in the uniform condition were also more likely than chance to accept V2-novel sentences compared to learners in the object-dominant

[8]The acceptance rate for V2-novel sentences in the object-dominant condition was in fact significantly below chance ($\beta = -1.75$, SE $= .41$, $p = 2.17 \times 10^{-5}$).

condition ($\beta = 1.08$, SE $= .46$, $p = .02$). The adjunct-dominant condition and the uniform condition were directly compared by fitting another model with identical effect structure to the same data with the adjunct-dominant condition as baseline for CONDITION. The model indicated that learners in the uniform language were not more likely to accept V2-novel sentences than learners in the adjunct-dominant language ($\beta = -0.08$, SE $= .45$, $p = .85$). The model thus confirmed my prediction for the contrast between the uniform condition and the object-dominant condition. Crucially, however, no differences were found for the contrast between the uniform condition and the adjunct-dominant condition.

The second prediction which concerns the discrimination between grammatical V2-novel and ungrammatical V3 sentences, was evaluated by examining the interaction between CONDITION and SENTENCE TYPE. V3 sentences were equally likely to be accepted as V2-novel sentences by participants in the object-dominant condition ($\beta = -0.15$, SE $= .48$, $p = .75$). The significant interaction between V3 and the adjunct-dominant condition ($\beta = -1.31$, SE $= .52$, $p = .01$) points towards a better discrimination between V2-novel and V3 sentences by participants in the adjunct-dominant condition. The interaction for V3 and the uniform condition ($\beta = -1.20$, SE $= .52$, $p = .02$) suggests that the uniform condition was also better at discriminating V2-novel and V3 sentences. The direct comparison of the adjunct-dominant condition with the uniform condition ($\beta = .12$, SE $= .51$, $p = .82$) did not provide evidence for a better discrimination by one of the conditions. In sum, my prediction was confirmed for the contrast between the uniform condition and the object-dominant condition. However, contrary to my prediction, no differences could be found between the uniform condition and the adjunct-dominant condition.

4.4.3.2 Exploratory analysis

The analysis of the fronted constituent types in participants' productions was based on icons that were lexically novel for participants. That is, participants had not previously seen these icons during training. As in experiment 4, the narrow focus was chosen to rule out the possibility that participants placed constituents in initial position because they were already seen in that position during training. The exploratory analysis of experiment 4 (§4.3.3.2), however, suggested that participants were not sensitive to the lexical novelty of novel constituents. I conducted an exploratory analysis to determine whether this was also the case for experiment 5. Figure 4.16 illustrates the proportion of lexically familiar and lexically novel fronted indirect objects. Two questions were addressed in the analysis, namely whether conditions differ with respect to the fronting of lexically familiar novel constituents and whether differences exist between lexically novel and lexically familiar novel constituents in terms of preposing to the clause-initial position. A mixed-effects logistic regression model was fitted to all produced V2 sentences. The dependent variable was the nature of the clause-initial constituent, i.e. whether a novel constituent type ($= 1$) or a familiar constituent type ($= 0$) was placed in initial position. The model included CONDITION and LEXICAL NOVELTY as fixed effects, as well as an interaction term for both. The model included by-participant random intercepts and by-participant random slopes for LEXICAL NOVELTY. By-item

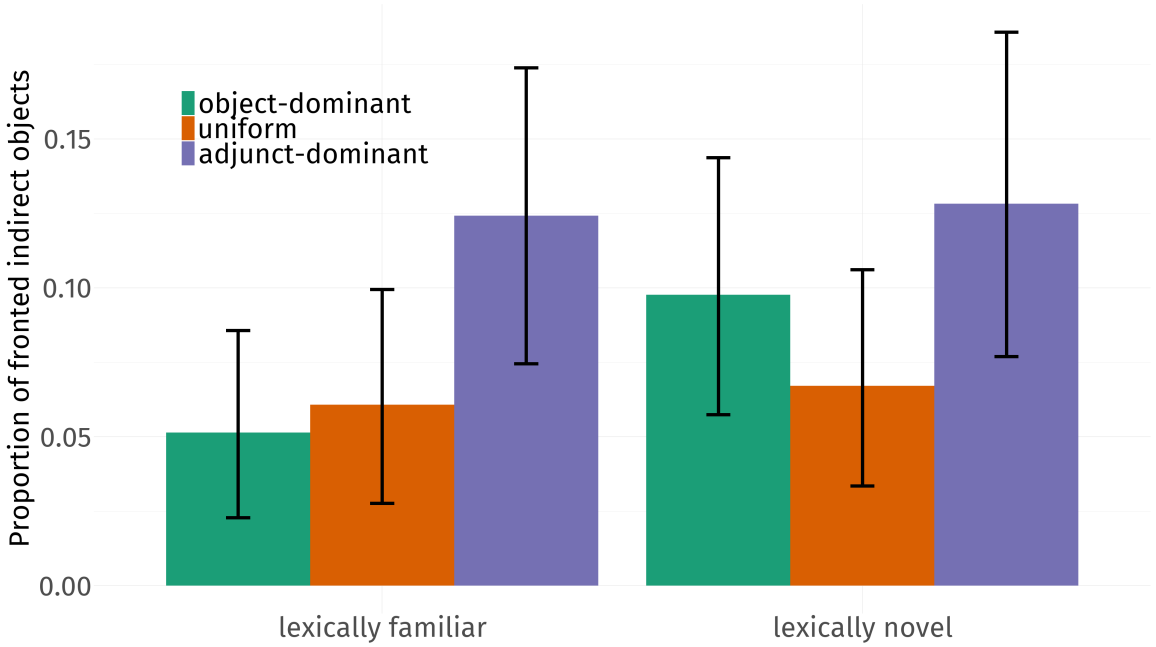


Figure 4.16: Proportion of lexically familiar and lexically novel indirect objects produced in clause-initial position in V2 sentences by participants in experiment 5. Error bars indicate bootstrapped 95% confidence intervals.

random intercepts were dropped due to singular fit. CONDITION and LEXICAL NOVELTY were treatment coded with object-dominant and lexically familiar as baseline. The first question, i.e. whether some conditions were more likely to place lexically familiar novel constituents in initial position, can be investigated by examining the simple effects of LEXICAL NOVELTY. The model indicated that participants in adjunct-dominant condition ($\beta = 1.40$, SE $= 1.24$, $p = .26$) and the uniform condition ($\beta = .30$, SE $= 1.26$, $p = .81$) were equally likely to place novel constituents in initial position as participants in the object-dominant condition. The second question concerned the differences between lexically familiar and lexically novel clause-initial novel constituents. Participants in the object-dominant condition were equally likely to produce lexically novel and lexically familiar novel constituents types in clause-initial position ($\beta = 1.60$, SE $= 1.34$, $p = .23$). The non-significant interactions between lexically novel and the adjunct-dominant condition ($\beta = -0.89$, SE $= .86$, $p = .30$) as well as the uniform condition ($\beta = -1.02$, SE $= .88$, $p = .25$) indicate that all conditions were equally likely to place lexically novel and lexically familiar novel constituent types in initial position.

In experiment 3, all conditions were equally likely to accept V2-familiar sentences. That is, the different training distributions affected only how well participants generalised, but not how well they learnt the language in the input. To examine whether this was also the case in experiment 5, I conducted a further exploratory analysis. Figure 4.15 visualises the acceptance rates for V2-familiar sentences in the present experiment.

The plot shows that the acceptance rates for V2-familiar was generally high across all conditions. A mixed-effects logistic regression model was fitted to all V2-familiar sentences. The model included CONDITION as fixed effect as well as by-participant and by-item random intercepts. CONDITION was treatment coded with the object-dominant condition as the baseline. Similar to experiment 3, the adjunct-dominant condition did not differ significantly from the object-dominant condition ($\beta = .03$, SE $= .28$, $p = .91$). The same pattern was observed when the uniform condition was compared with the object-dominant condition ($\beta = .14$, SE $= .29$, $p = .63$). These results thus suggest that participants in all three conditions learnt the language they were trained on equally well.

The analysis of V2 and V3 sentences with different initial constituent types in experiment 4 has shown that participants struggled to discriminate between grammatical and ungrammatical object-initial and adjunct-initial sentences. This was surprising as participants were previously trained on these types. If participants had learnt the language, they should have been able to distinguish between grammatical V2 and ungrammatical V3 sentences. The only constituent type for which participants could reliably distinguish grammatical and ungrammatical sentences were initial subjects. This finding also proved crucial for the interpretation of the acceptance rates for V2-familiar sentences: The high acceptance rates for V2-familiar sentences were substantially driven by the very high acceptance rates of subject-initial V2 sentences. This finding calls for a closer examination of the acceptance rates of V2 and V3 sentences with different clause-initial constituents obtained in experiment 5. Figure 4.17 shows the acceptance rate for different clause-initial constituents. The plot shows that, akin to experiment 4, participants excelled at correctly identifying grammatical and ungrammatical subject-initial test sentences. To examine whether participants were also able to discriminate between adjunct-initial and object-initial V2 and V3 sentences, two mixed-effects logistic regression models were fitted to adjunct-initial sentences and object-initial sentences, respectively. The models included CONDITION and WORD ORDER as fixed effects as well as an interaction term for both. The model further comprised by-participant and by-item random intercepts as well as by-participant random slopes for WORD ORDER. Both fixed effects were treatment coded with object-dominant and V2 as baseline. I examined first the acceptance rates for adjunct-initial sentences. The intercept of the model suggests that the acceptance rate for adjunct-initial V2 sentences were rated significantly above chance ($\beta = 1.38$, SE $= .46$, $p = .002$). The model further indicated that all conditions were equally likely to accept adjunct-initial V2 sentences (adjunct-dominant: $\beta = .97$, SE $= .60$, $p = .11$; uniform: $\beta = .29$, SE $= .59$, $p = .63$). Participants in the object-dominant condition were as likely to accept V3 sentences as V2 sentences ($\beta = -1.28$, SE $= .87$, $p = .14$). The non-significant interactions of V3 and the adjunct-dominant condition ($\beta = -2.08$, SE $= 1.20$, $p = .08$) show that participants in this condition were not better at discriminating V2 and V3 sentences than participants in the object-dominant condition. Similarly for the uniform condition, the non-significant interaction of V3 and the uniform condition ($\beta = -0.26$, SE $=$ 1.19, $p = .83$) also suggests that participants in this condition were not more likely to discriminate between grammatical and ungrammatical adjunct-initial sentences. In

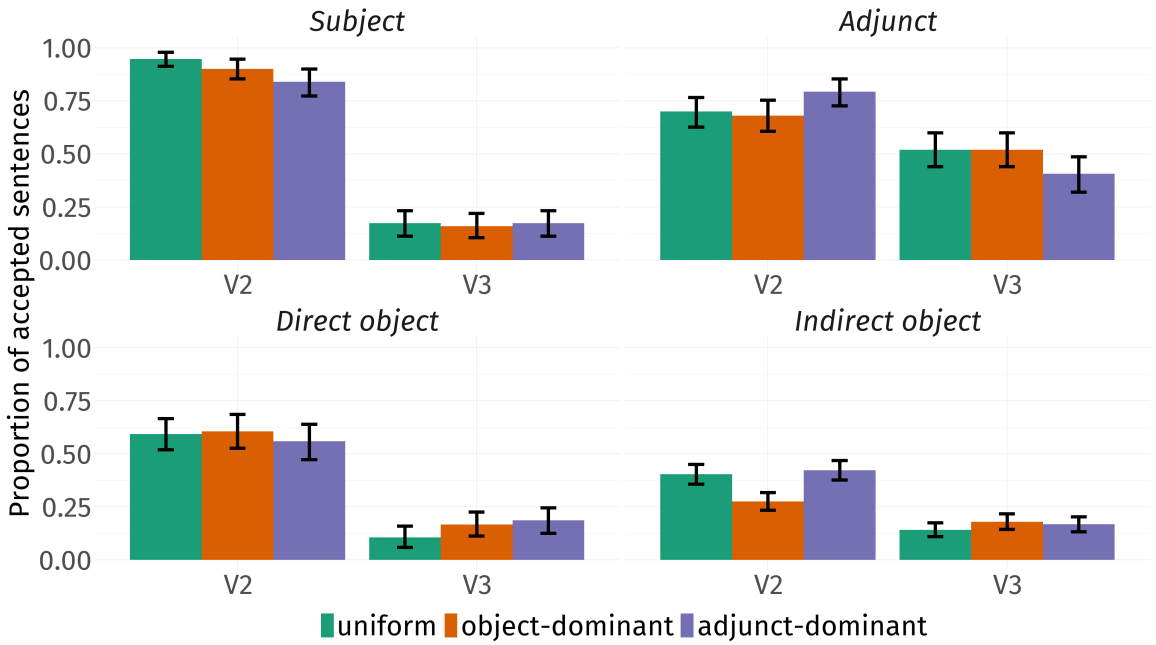


Figure 4.17: Acceptance rates of V2 and V3 sentences with different clause-initial constituents by condition in the judgement task in experiment 5. Error bars indicate bootstrapped 95% confidence intervals. Subjects, direct objects and adjuncts were familiar from training. Participants were not exposed to indirect objects during training, they thus constitute a novel constituent type. The ratings for indirect objects are included for reference only.

sum, participants across all conditions failed to discriminate between grammatical V2 and ungrammatical V3 sentences with adjuncts in initial position.

Next, the judgements for the object-initial sentences were scrutinised more closely. The intercept of the model revealed that participants in the object-dominant condition rated V2 sentences with initial objects significantly above chance ($\beta = .80$, SE $= .40$, $p = .047$). The adjunct-dominant condition ($\beta = -.38$, SE $= .55$, $p = .49$) and the uniform condition ($\beta = -0.14$, SE $= .55$, $p = .80$) were as likely to accept object-initial V2 sentences as the object-dominant condition. The model showed further that participants in the object-dominant condition were significantly less likely to accept ungrammatical object-initial V3 sentences than grammatical V2 sentences ($\beta = -6.41$, SE $= 1.59$, $p = 5.49 \times 10^{-5}$). The non-significant interactions of V3 and the adjunct-dominant condition ($\beta = .84$, SE $= 1.12$, $p = .45$) and the uniform condition ($\beta = -0.54$, SE $= 1.13$, $p = .64$) suggest that none of the other two conditions showed a better or worse discrimination between grammatical and ungrammatical object-initial sentences. That is, participants across all conditions were able to discriminate between grammatical and ungrammatical object-initial sentences in their judgements.

4.4.4 Discussion

The purpose of experiment 4 was to replicate the findings of experiment 3 (§3.2) with a language that exhibits fewer similarities to participants' native language (i.e. English). This was motivated by the desire to better understand whether the findings of experiment 3 — better learning of an adjunct-dominant language, and worst learning of an object-dominant language — were driven, at least in part, by transfer from English. We therefore attempted to reduce the similarities between the artificial language and English; less interference should follow. However, the failure of learners to robustly acquire a V2 grammar in experiment 4, potentially due to the lack of variability, prompted me to consider an alternative to 'conventional' ALL: iconic artificial language learning (Shapiro & Steinert-Threlkeld 2023). In this experimental paradigm, words are replaced by icons representing the same concepts as the words. Because the language is iconic, there are two important advantages of this paradigm in this case. On the one hand, a much larger lexicon can be used in a single experimental session. On the other hand, participants can focus on the syntactic patterns of interest. That is, iconic artificial language learning offers the same advantages as semi-artificial languages without the potential confound of significant interference from the L1. I therefore adopted this paradigm to repeat experiment 4 as iconic artificial language learning experiment in experiment 5. The design remained the same in that different distributions of clause-initial elements were compared: a uniform distribution and two skewed distributions, namely an object-dominant distribution and an adjunct-dominant distribution. The subject-dominant distribution used in experiment 4 was omitted due to the strong SVO bias observed there. Learning was again measured as extrapolation to novel structures in productions and judgements.

Irrespective of the condition, participants in experiment 5 learnt to place the verb in second position. Crucially, no differences between learners in the three conditions were observed. This was expected given that the verb was consistently placed in second position during training. More important is thus the question of whether participants also learnt that different constituents can occupy the initial position. Participants in experiment 4 failed to learn the language they were trained on. This was reflected by their general inability to discriminate between grammatical and ungrammatical sentences with familiar constituent types in initial position. The production data from experiment 5 does not provide strong evidence that the language was learnt, given the reluctance to place non-subjects in initial position. However, the judgement data provided clearer evidence that learners in experiment 5 indeed acquired the language they were trained on. Object-initial V2 sentences were more likely to be accepted than their V3 counterparts. Moreover, the significantly above chance ratings of object-initial V2 sentences by participants in the object-dominant condition and absence of any differences between the object-dominant condition and the other two conditions show that participants have developed knowledge of the language.

The ratings for adjunct-initial V2 sentences also suggest that learners correctly recognised these structures. However, participants struggled to differentiate between V2 and V3 sentences of this type in their judgements. Although this could be interpreted as the failure of participants to learn some aspects of the input language, this is not

the only explanation. First, if participants did not learn that adjunct-initial sentences are licensed by the grammar, the high acceptance rates for V2 sentences in all three conditions are unexpected. Second, the acceptance rates for adjunct-initial V3 sentences are the only type of V3 sentences that received such high ratings (cf. Figure 4.17). This suggests that adjunct-initial V3 sentences may have been treated as special, potentially because they are possible in English. As mentioned in §3.2, topicalisations in English are particularly common with adjuncts (Doherty 2005). That is, participants could have transferred structures like (2) from English into the visual language. This transfer could then explain the high acceptance rate.

(2) In the cheese, *the diver* **found** the mouse sleeping.

An alternative explanation is cross-linguistic in nature. The adjuncts used in experiment 5 could function as scene-setters when realised in initial position. That is, they provide the context in which the following utterance is interpreted. Cross-linguistically, scene-setters tend to be realised towards the beginning of a sentence (Benincà & Poletto 2004, Napoli & Sutton–Spence 2021), as the Italian example in (3) demonstrates. There, *domani* 'tomorrow' sets the 'scene' for the proposition, i.e. the subject meeting Gianni. This fact is also reflected in syntactic analyses assuming an articulated left periphery (cf. (38) in Chapter 1; Benincà & Poletto 2004). If this cross-linguistic pattern is motivated by properties of human cognition (cf. Culbertson, Smolensky & Legendre 2012, Culbertson et al. 2020), this could then explain the preference for V3 sentences with initial frame-setters as opposed to V3 sentences with other initial constituent types.[9] Wiese et al. (2020) argue in a similar direction in that the language-independent pattern of initial scene-setters is attributed to general information-structural preferences.

(3) Domani Gianni lo **vedo**. ITALIAN
tomorrow Gianni him see
'Tomorrow I will see Gianni.'
(Benincà & Poletto 2004: 67)

Interestingly, deviations from V2 word orders (in the Germanic V2 languages at least) follow the exact same pattern, namely scene-setter > subject > verb. Although this has been mostly observed for urban vernaculars such as *Kiezdeutsch* '(lit.) hood German' (Freywald et al. 2015, Walkden 2017a, Meelen, Mourigh & Cheng 2020), similar orders have been noted for West Flemish (Greco & Haegeman 2020) and historic stages of English (cf. §1.3.1) and French (cf. §1.3.2). If participants learnt a V2 language, high ratings for adjunct-initial V3 sentences would be unsurprising from this perspective. To summarise, the high acceptance rates of adjunct-initial V3 sentences does *not* necessarily argue against the conclusion that participants in experiment 5 learnt the language they were trained on.

[9] Such an explanation is indeed plausible. Bross (2020), for instance, argues that scope in German Sign Language (DGS) correlates with the position of the articulator in the sense that "the higher the scope the higher the articulator" (Bross 2020: 23). In spoken language, this could translate to a position early in the clause. Scene-setters can take scope over the whole sentence and should thus appear first.

While no differences were observed for V2 sentences with familiar initial constituent types, learners displayed different learning patterns when it comes to generalising XP-fronting. As predicted, participants in the uniform condition were more likely to accept V2 sentences with novel constituent types (i.e. indirect objects) in initial position than participants in the object-dominant condition. Participants in the adjunct-dominant condition were as likely to accept V2 sentences with novel clause-initial constituents as those in the uniform condition. The same pattern was observed for the discrimination between grammatical V2 and ungrammatical V3 sentences. Learners in the object-dominant condition failed to discriminate between the two sentence types, whereas learners of the uniform and adjunct-dominant condition did. Again, no difference in the ability to discriminate could be found between the uniform condition and the adjunct-dominant condition. The production data did not reveal a difference between the conditions. All three conditions were equally likely to place novel constituent types in clause-initial position, irrespective of the lexical novelty of the constituents. Learners were generally hesitant to produce V2 sentences with non-subjects in initial position. This parallels the production data in experiment 4. Taken together, learners in the uniform condition and the adjunct-dominant condition exhibited a learning advantage over learners in the object-dominant condition. In other words, only learners in the object-dominant condition were unable to extrapolate to novel contexts, as the failure to discriminate between grammatical and ungrammatical sentences suggests.

The results of experiment 5 differ significantly from those in experiment 4. On the one hand, I evidence that participants across all three conditions were able to learn the language in the input. On the other hand, participants in the uniform condition and the adjunct-dominant condition were able to extrapolate XP-fronting to novel contexts. This suggests that the larger lexical variability (and potentially the increased number of training trials) benefited learning. The present findings therefore provide further evidence that encountering additional variability during learning supports the learning process. In addition, experiment 5 has contributed to the development of a new iconic artificial language learning; it provides further support that this paradigm constitutes a suitable alternative for researchers in syntax.

Before the insights gained in experiment 5 can be discussed in the context of experiment 3, the extent to which participants' native language might have affected the results has to be evaluated. The underlying assumption for the use of both the fully-artificial language and the iconic artificial language was a presumed weaker influence of participants' L1. Shapiro & Steinert-Threlkeld (2023: 151) report that many of their participants verbalised icons which may have triggered transfer from participants' native language. I therefore reviewed the post-test questionnaire responses in experiment 5. While many participants did not indicate to have used any strategies, six participants alluded to the use of English to complete the experiment. This use may either be in the form of (silently) reading the icons or reference to English grammar. Considering the fact that over 150 participants were tested, this number is very low. That English does not, at least explicitly, drive participants' behaviour in iconic artificial language learning experiments is also supported by the analysis of Shapiro & Steinert-Threlkeld (2023: 151). Although many participants in that study verbalised icons, the transfer

effects were much less pronounced than in the studies of Culbertson & Adger (2014) and Martin et al. (2019). That is, even if participants articulated icons in my experiment, the transfer effect may still be lower than with a semi-artificial language. In consequence, I argue that the effects are more likely the result of the learning distribution manipulated in the experiment, plus potentially more general preferences for ordering of particular elements. Against the backdrop of this conclusion, the results of experiments 3 and 5 can be compared.

4.5 General discussion

In this section, I focus on comparing the results of experiment 3 and 5 — two experiments in which participants show clear evidence of having learnt the input language they were trained on, and some evidence for extrapolation to novel clause-initial elements. In both experiments, learners in all conditions learnt the input language equally well, as reflected in the judgements for V2 sentences with familiar initial constituent types. Second, learners in the uniform condition were better at generalising V2 to novel contexts than learners in the object-dominant condition. The same applies to the comparison of learners in the adjunct-dominant condition with the object-dominant condition. In both experiments, learners in the adjunct-dominant condition also exhibited a learning advantage over those in the object-dominant condition. At the same time, however, there were differences between the two experiments. The learning advantages observed for the uniform condition and the adjunct-dominant condition over the object-dominant condition did not extend to production in experiment 5. Across conditions, a strong preference for SVO sentences was observable. Moreover, the learning advantage of the adjunct-dominant condition over the uniform condition — both in terms of production and judgement in experiment 3 — was not replicated in experiment 5. Both groups of learners showed similar behaviours. A final difference is the extent to which participants extrapolated. A comparison of the acceptance rates for V2-novel sentences in experiment 3 (Figure 3.5) and experiment 5 (Figure 4.15) reveals considerable differences. While the mean acceptance rates for V2-novel sentences in experiment 3 were over 80% in the adjunct-dominant condition and over 60% in the uniform condition, the mean acceptance rates in experiment 5 are around 40% in both conditions. That is, a considerable gap exists. In what follows, I will first focus on the differences in the results of experiment 3 and experiment 5, before an analysis is proposed that can explain the differences and similarities.

Although the results for the production data differ across experiments, these differences may not be as clear-cut as they appear at first sight. When the adjunct-dominant condition in experiment 3 is left aside, the extent to which participants placed novel constituent types in initial position is comparable. In the uniform condition, novel constituents were realised in initial position in 14% and 7% of the produced V2 sentences in experiment 3 and 5, respectively. Learners in the object-dominant condition exhibited a similar behaviour in that 8% in experiment 3 and 10% in experiment 5 of all constituents in V2 sentences were novel types. That is, in both experiments learners were generally hesitant to use novel constituents. Learners in the adjunct-dominant

condition in experiment 3 can thus be conceived as outliers given the high amount of novel constituents (47% as opposed to 13% in experiment 5). This poses two questions. First, why do learners in the adjunct-dominant condition in experiment 3 diverge so significantly from those in the other conditions? And second, why are participants hesitant to place novel types of constituents in clause-initial position? A possible explanation for these pattern could be that learners in the adjunct-dominant condition picked up on a similarity between adjuncts and indirect objects: both are realised as PPs. All indirect objects in experiment 3 were realised as prepositional objects (e.g. *to the congregation*). The formal identity to adjuncts would thus facilitate such an interpretation. As learners were familiar with adjuncts in initial position, they might have been more likely to place indirect objects in initial position than learners in other conditions. By contrast, learners in experiment 5 do not have the same evidence for a syntactic similarity between indirect objects as adjuncts — they were distinct in the types of meanings they conveyed (person versus environment) and there was no evidence that they were any more syntactically similar than any other category expressed in the language.

As for the second question — i.e. why learners refrained from realising non-subjects in initial position — only a tentative interpretation can be provided. As discussed previously (e.g. §3.2), corpus evidence suggests that in V2 languages, objects account only for a small minority of clause-initial constituents. My corpus study also revealed that indirect objects occur even less frequently clause-initially than direct objects. The fact that non-subjects were only very rarely produced in clause-initial position thus aligns with data from natural V2 languages. Object-fronting in German and Swedish has been argued to be subject to information-structural constraints. According to Holmberg (2015: 348), objects can be placed in clause-initial position in Swedish if they are (contrastive) topics. Similarly in German, preposed objects tend to be given but also interact with the nature of the subject (Bader 2020, Bader & Portele 2021). If learners posited similar constraints on the information-structural status of fronted objects, the low number of fronted objects could be explained.[10,11]

A further difference between experiments 3 and 5, as noted above, is the degree of extrapolation in participants' judgements. Learners in the uniform condition and the adjunct-dominant condition were more likely to extrapolate XP-fronting to novel constituents in experiment 3 compared to experiment 5. This is likely a task-related effect. The post-test questionnaire in experiment 5 included a question specifically inquiring about the strategies participants used to judge sentences with two persons — in other words sentences that included an indirect object. Several participants in experiment 5 indicated that these specific sentences were difficult to judge. For instance,

[10] Admittedly, this explanation leaves many questions open. For instance, why should learners assume such constraints on the information-structural organisation?

[11] Initially, one might be tempted to attribute the lack of objects in clause-initial position to more general language processing or cognitive biases. Animate words have been shown to be generally preferred and they are retrieved and processed faster than inanimate words (Branigan, Pickering & Tanaka 2008, de Swart & van Bergen 2019); a number of studies using improvised silent gesture have found a dispreference for initial-objects, particularly when they are inanimate, mirroring natural language typology (e.g. Goldin-Meadow et al. 2008, among many others). The rarity of initial animate indirect objects does not necessarily follow from these findings though.

one participant reported that two people in the same sentence were hard to decode. This suggests participants in experiment 5 may have had more difficulty identifying the subject and the indirect object compared to those in experiment 3. It appears as if the initial position was by default interpreted as subject when the constituent was animate. In cases where the images illustrating the sentence did not match this interpretation, sentences were more likely to be rejected. If at least some participants experienced similar issues, as the questionnaire responses indeed suggest, the lower extrapolation rate can be interpreted as the direct result of this. That is, the lower rates are likely the results of the task design. In particular, as noted above, in experiment 5 there was less evidence for the syntactic status of indirect objects compared to experiment 3.

I now turn to the third and final difference between both experiments, i.e. the absence of a learning advantage of the adjunct-dominant condition over the uniform condition in experiment 5. This finding does not match the explanation proposed for experiment 3: that a large proportion of initial adjuncts fosters learning. There, the learning advantage was attributed to a more uniform distribution in grammatical categories (e.g. NP, PP). The discussion in §3.3 has shown that learners in the adjunct-dominant condition were exposed to the highest variability in clause-initial grammatical categories. Learners in the uniform condition, in turn, experienced less variability in grammatical categories which, however, was still higher than the one in the object-dominant condition. From this perspective, the absence of a learning advantage in experiment 5 is particularly striking since the distribution of grammatical categories in this experiment was actually perfectly uniform in the adjunct-dominant condition. Recall that the dominating element accounted for 50% (instead of 60% as in experiment 4) of all initial constituents during training. Consequently, half of the clause-initial constituent types were PPs (and/or related types),[12] the other half were NPs (25% subjects, 25% objects). This means that the variability in the grammatical categories was at its maximum. As a result, both adjunct-dominant conditions exhibited the highest variability in grammatical functions. At the same time, the results also do not align with the prediction that more variability in the grammatical functions in clause-initial position leads to better learning of a V2 grammar. The hypothesis would be borne out if learners in the uniform condition exhibited a learning advantage over learners in the adjunct-dominant condition. One intriguing interpretation that would allow me to provide a unifying explanation for the results of experiment 3 and 5 while maintaining the role of variability is the following: Learners can be sensitive to both types of variability when learning V2, that is variability in grammatical functions and variability in grammatical categories; the relevant type of variability is determined by the context: If learners encounter high variability in grammatical functions, they focus on this aspect. If, however, learners are faced with high variability in grammatical categories, they will devote their attention to that property. A V2 grammar might thus be formulated either as (4) following Yang (2000, 2002), or as (5) following Lightfoot (1999, 2006). Since syntactic rules in general can be phrased with either notion (cf. §1.5.2), no option should be ruled out *a priori*.

(4) Any grammatical *function* can reside in the clause-initial position in a V2 clause.

[12] It is difficult to determine which grammatical category the adjunct icons exactly correspond to.

(5) Any grammatical *category* can reside in the clause-initial position in a V2 clause.

The analysis suggested here can explain why learners in the uniform condition and the adjunct-dominant condition displayed similar behaviours in experiment 5 with respect to generalising. The former group derived (4) from the input, the latter group (5). Furthermore, irrespective of the relevant variability domain, the object-dominant condition always exhibited the least amount of variability. The observed patterns of learners in the object-dominant condition are therefore also predicted.

How can the learning advantage of the adjunct-dominant condition over the uniform language in experiment 3 then be explained? Above, I argued that the influence of English (i.e. the L1) was less pronounced in experiment 5. Conversely, this must entail a larger influence of English on the learning in experiment 3. Such influence would provide support for one of my original interpretations of the results, namely that participants' L1 in the adjunct-dominant condition facilitated the acquisition of the V2 grammar. If both variability and L1 influence are at play, then all of the patterns I have noted here can be explained: Learners in the adjunct-dominant condition of experiment 3 received a two-fold boost — variability plus L1 interference — for learning the V2 grammar and, as a result, performed best. Learners in the uniform condition only profited from the variability in the initial position and thus performed better than the learners in the object-dominant condition who did not benefit at all from their training distribution (and may have be disadvantaged by a general dispreference for initial objects). Note that the results of the corpus study are also compatible with this analysis. A high proportion of adjuncts results either in more variability in terms of grammatical functions or grammatical categories. The difference in the results between experiment 3 and experiment 5 can thus be attributed to the change in the experiment design, that is the switch from a semi-artificial language to an iconic artificial language. It can, of course, not be completely ruled out that changes to the input frequencies during training also affected the results. Crucially however, the frequencies of DPs were almost similar in the adjunct-dominant conditions of experiment 3 (52%, cf. §3.3) and experiment 5 (50%). Moreover, even though adjuncts were realised as both AdvPs and PPs in experiment 3, the vast majority of adjuncts were AdvPs (40%). Hence, just two types of grammatical categories account for at least 90% of the training data in both experiments.[13] The main point of divergence thus lies in the nature of the language.

Taken together, the preceding discussion suggests that the results of experiment 3 were indeed affected by the native language of the participants in that experiment. However, the results of experiment 5 also indicate that L1 interference is not solely responsible for the observed patterns. Variability indeed fosters learning, but the nature of that variability may depend on the context. Learners might make notice of variability

[13]According to my hypothesis, the larger variability of the clause-initial position in experiment 3 should lead to a better learning outcome compared to experiment 5. The entropy of the clause-initial position (which quantifies variability) in the adjunct-dominant dominant condition was 1.31 bits in experiment 3 and 1 bit in experiment 5. The experimental results do not allow me to draw any conclusion in this regard. The possibility of a difference, however, must be entertained in future work.

in either functions or categories; and both might thus be beneficial for the acquisition of a V2 grammar.

4.6 Summary

The goal of the present chapter was to distinguish between the two interpretations proposed for the results of experiment 3 in Chapter 3: (i) interference from learners' native language and (ii) a fostering effect of a large proportion of adjuncts in initial position. To reduce the presumed effect of learners' L1, the experiment was repeated using a fully-artificial language, while relying on an otherwise similar design as in experiment 3. The analysis showed however that participants across all conditions were unable to learn the language they were trained on, let alone extrapolate V2 to novel structures. The lack of sufficient variability in the lexicon, in line with the overall hypothesis of this thesis, was suggested as reason for the null result in experiment 4. The experiment was therefore repeated with an iconic artificial language in which words were replaced by icons. This type of language offers the advantage that a relatively large variety of lexical items can be used without any prior training. Learners in all three conditions (i.e. uniform, object-dominant and adjunct-dominant) were able to acquire the language they were trained on. Yet, only learners in the uniform condition and the adjunct-dominant condition generalised V2. In addition, the degree to which participants generalised is similar in both conditions. The generalisation of V2 was only observable in the judgements that learners provided. The results suggest that the learning advantage of learners in experiment 3 was partially the result of interference from learners' native language. However, the results also showed that higher variability in the clause-initial position does indeed foster the acquisition of a V2 language. The lack of a difference between learners in the uniform and adjunct-dominant condition further indicates that participants can be sensitive to variability in initial grammatical functions (e.g. subject, object) and variability in initial grammatical categories (e.g. NP, PP). The implications of these findings will be discussed in Chapter 5.

CHAPTER 5

GENERAL DISCUSSION

5.1 Summary of findings

This thesis investigated how variability in the constituent types realised in the clause-initial position of V2 sentences affects the acquisition of a V2 grammar. The research question was motivated by two observations made in Chapter 1. First, languages that have lost V2 show a similar development in that subject-initial V2 sentences increase, whereas non-subject-initial V2 sentences decrease. Second, variability in the input fosters language acquisition. This led to the assumption that the decrease of non-subjects in initial position in V2 sentences might play a role in the loss of V2. If such a connection between variability in the clause-initial position and learning a V2 grammar was established, general properties of human cognition could be determined as one of the factors contributing to the loss of V2 across languages. Specifically, I hypothesised that the lack of variability in the clause-initial position will lead to the loss of V2, because learners are unable to form generalisations about the flexibility of the clause-initial position. Instead, learners will posit a fixed association of the clause-initial position with a particular grammatical property. Furthermore, I hypothesised that the relevant domain of variability is grammatical functions (e.g. subjects, direct objects) and not grammatical categories (e.g. NP, PP). Note that this deviates from previous studies relating the loss of V2 to learning in that those studies focused on unambiguous evidence for a V2 grammar (Lightfoot 1999, 2006, Yang 2000, 2002, Westergaard 2008, 2009b). In order to test my hypothesis, I developed an experimental design using artificial language learning (ALL) in Chapter 2. By using artificial languages, the variability in the clause-initial position can easily be manipulated, thus enabling me to directly test the hypothesis. Building on insights from previous work by Rebuschat (2008) and colleagues as well as Getz (2018), I adopted a design relying on a semi-artificial language with V2 syntax and English vocabulary after testing (experiments 1 & 2). An important innovation of the design was the introduction of tests to measure how well learners generalise V2 to novel contexts. Such tests can be used to assess whether participants have actually learnt a V2 language as opposed to a language that licenses a limited number of constituent types in initial position.

Chapter 3 presented the results of an ALL experiment (experiment 3) that put the hypothesis investigated here to test using the design developed in Chapter 2. I compared the effects on learning of three different distributions of clause-initial constituents: a uniform language in which subjects, direct objects and adjuncts occurred with the

same frequency in clause-initial position and two languages with skewed distributions with either objects or adjuncts as the most frequent initial element. According to the hypothesis from Chapter 1, the uniform language should be learnt best as it featured the highest variability. The results, however, showed that learners of the adjunct-dominant language performed best at generalising V2 to novel contexts, and learners of the object-dominant language performed worst. That is, the hypothesis only correctly predicted the difference between the uniform language and the object-dominant language. Crucially, no difference between the languages was found for the learning of structures familiar from training. I offered two (mutually non-exclusive) interpretations. First, learners' native language could have given those learning the adjunct-dominant language an advantage. Second, a large proportion of initial adjuncts might be beneficial for learning more generally. This could be explained if learners were sensitive to variability in the grammatical categories rather than grammatical functions as I original assumed. To support the second interpretation, I conducted a large-scale corpus study investigating the distribution of clause-initial elements in German: Although subjects account for the most frequent initial element, adjuncts are the second most frequent type. This finding also converges with other studies on German and other V2 languages reported in the literature that examined a smaller sample.

To distinguish between the two explanations proposed in Chapter 3, I conducted two further experiments and reported them in Chapter 4. Following previous work that tried to replicate findings of experiments with semi-artificial languages, I first used a fully-artificial language (experiment 4). In contrast to all previous experiments, the language was not learnt by participants. The analysis suggested that participants might have analysed the languages as SVO instead. I attributed the failure to learn the language to a lack of variability, in line with the overall hypothesis investigated in this thesis: the lexicon size was too small and the number of training items too low to learn a V2 language. To counter these issues, I adopted a novel paradigm in a follow-up experiment (experiment 5) — iconic artificial language learning (Shapiro & Steinert-Threlkeld 2023). In this type of design, lexical items are replaced by icons. The meaning of icons are generally easily identifiable and do not necessitate a special training session. As a result, the lexicon size of the language could be significantly increased. Unlike the fully-artificial language, the iconic artificial language could indeed be learnt. The results revealed that learners of an adjunct-dominant language and a uniform language performed better at test than those learning an object-dominant condition. Importantly, I found no difference between learners of the adjunct-dominant language and the uniform language. As in the experiment with the semi-artificial language, all learners learnt the structures from training equally well irrespective of the distribution. A comparison of the experiment with the iconic artificial language and the experiment with the semi-artificial language suggested that the learning advantage of the adjunct-dominant language over the uniform language in the latter experiment was likely the result of L1 interference. However, the persisting learning advantage of the adjunct-dominant language and the uniform language over the object-dominant language does indeed suggest that variability fosters learning. Crucially, the learning

advantage can arise as a result of variability in grammatical functions or in grammatical categories.

5.2 Implications

The findings reported in this thesis have implications for both the analysis of the loss of V2 and language change more generally, but also for the methodological advancement of ALL. These implications will be discussed in §5.2.1 and §5.2.2, respectively.

5.2.1 Loss of V2

The goal of this thesis is to explore whether loss of variability in initial constituents — a development that has occurred in all of the (known) languages to have lost V2 — can be explained by features of human cognition that impact learning. Importantly, of course, even if this is the case, it is very likely not the only reason V2 might have been lost in these languages. Indeed, many other factors have been proposed, which have not been investigated here, such as the rise of V3 constructions. A learning perspective does not constitute a novel approach in and of itself though — existing analyses of the loss of V2 in individual languages have incorporated learning (albeit only vaguely in some cases). Furthermore, learning models such as the variational learning model Yang (2000, 2002) or cue-based approaches in different guises (Lightfoot 1999, 2006, Westergaard 2008, 2009b) emphasise the role of learning in the loss of V2. However, unlike these learning models, which underscore the role of ambiguous evidence (subject-initial V2 sentences) versus unambiguous evidence (non-subject-initial V2 sentences) for V2, the approach adopted here focuses on a different aspect of the input, namely its composition. Based on a domain-general fostering effect of variability (cf. Raviv, Lupyan & Green 2022), I hypothesised that low variability in the clause-initial position will lead to the loss of V2 as learners are unable to form generalisations about the flexibility of the clause-initial position. Instead, learners will stipulate a fixed association of the clause-initial position with a particular grammatical property.

The diachronic developments noted in §1.3 support this view: Language-specific developments, mostly extraneous to the syntactic domain, lead to a redistribution of clause-initial elements in V2 clauses. As a result, the number of subject-initial sentences grows, at the cost of non-subject-initial sentences. One can therefore distinguish between proximal and distal causes for the loss of V2. Proximal causes encompass the immediate causes, such as changes in the variability of clause-initial constituents as assumed here. Distal causes, on the other hand, are those that bring about the proximal causes. That is, factors which lead to the redistribution of clause-initial elements. The present thesis therefore aimed at providing a uniform analysis of V2 in terms of proximal causes.

The experimental findings presented in this thesis provide substantial evidence for the view I advocate here. The experiments revealed learners' sensitivity to different distributions of clause-initial constituents. As predicted, V2 languages with higher initial variability were learnt better. That is, learners of a high-variable language were more likely to extrapolate V2 to novel contexts than learners of languages with low

clause-initial variability. Crucially, the effect of different distributions became only visible when learners had to extrapolate. Structures that learners were already familiar with from training (i.e. subject-initial, direct-object-initial and adjunct-initial sentences) were learnt equally well. This strongly suggests that the representations formed by learners of languages with high initial variability featured no (or at least less strong) fixed associations of the clause-initial position with particular constituent types. By contrast, learners of a low variable language associated the clause-initial position with particular constituent types, in the present case subjects, direct objects and adjuncts. In the light of these findings then, the lack of sufficient variability due to the rise of subject-initial V2 sentences can indeed contribute to the loss of the V2 property, at least in the languages discussed in Chapter 1 (i.e. English, French, Portuguese and Welsh). The fact that multiple languages showed similar developments despite not all being closely related suggests that the proposed analysis is a suitable account for explaining the loss of V2 more generally. However, future work examining the loss of V2 in other languages such as Italian or Spanish needs to confirm this. As for the distal causes of the loss of V2, the proposals made in the literature can be maintained.

Although only adults were tested in the experiments reported here, there is good reason to believe that this should not affect the interpretability of the results. Gómez (2002) and Gómez & Maye (2005) have demonstrated that adults and children show the same responses to variability in other areas of syntactic learning. Hence, it is reasonable to assume that the same holds for the current study.

The discussion so far steered clear of a specification of the relevant domain of variability in the clause-initial position. Although it was stipulated in §1.5.2 that variability in grammatical functions (e.g. subjects, direct objects) is decisive, this was purely based on the fact that grammars are typically defined in terms of grammatical functions (e.g. SOV, SVO). There was no inherent reason to assume that learners cannot be sensitive to variability in grammatical categories (e.g. NP, DP). This was confirmed by the findings of experiment 5 where learners displayed sensitivity to grammatical functions or grammatical categories, depending on the language they were learning. This then poses the question what the relevant domain of variability is in natural languages. Based on the current results, no definitive answer can be given and future work has to scrutinise this question more closely. It is conceivable, however, that in fact there is no general answer and the domain learners choose to focus on is contingent on their personal experience. For instance, one can imagine a situation in which one child is exposed to to more variability in grammatical functions in their input and therefore chooses to focus on this aspect of the input. Another child by contrast could be exposed to more variability in grammatical categories and thus focuses on this domain.

A final point to consider is the situation in the contemporary V2 languages. Even though the best learning outcome was achieved by learners of languages with the highest variability in the clause-initial position, this does not entail that a maximally variable language is actually necessary for the acquisition of V2. The distributions of clause-initial elements in natural V2 languages have been shown to be skewed: subjects are the most frequent type followed by adjuncts. This skew suggests that these languages

still exhibit sufficient variability for the language to be acquired.[1] The high proportion of adjuncts should also not be interpreted as evidence for a special status of adjuncts in the acquisition of V2. Irrespective of the relevant domain of variability (i.e. grammatical functions or grammatical categories) a high proportion of adjuncts will increase the variability in the clause-initial position.

Before I turn to the methodological implications of this thesis, it is worth asking in the context of the current findings who the instigators of the loss of V2 are. Two groups can be suspected to be the 'culprits'. Adults might be responsible for the change as they are likely the ones that alter the distribution which learners are then exposed to. This view would be compatible with previous proposals (e.g. Bybee & Slobin 1982, Diessel 2011). Alternatively, children could be conceived as the drivers of change since their learning response to the input causes the V2 system to become unstable. The second approach would be in line with the child innovator approach of Cournane (2017, 2019, cf. also Cournane & Klævik-Pettersen 2023) that connects syntactic change with language acquisition. Under my proposal, however, there is no need to single one group out as the sole driver of change. Instead, it is the interaction of both groups that brings about the changing patterns. As users, adults' productions can be subject to linguistic and extralinguistic influences. The latter might manifest themselves in the membership of particular socio-economic groups or the state of the immediate environment (e.g. a salient entity that is repeatedly under discussion). On the other hand, examples for the former type of influences could be found in the aforementioned distal causes of the loss of V2. Both, the linguistic and extralinguistic influences could determine the choice of clause-initial elements in V2 sentences, for instance due to the information-structural organisation of an utterance or the preference for a particular structure.[2] Adults constitute, at least initially, the main source of input for infants acquiring language. In the light of the findings of this thesis, shifts in the distribution of initial elements in V2 sentences would have a significant effect on the learning outcome. If the variability in the clause-initial constituent types declined, children would form representations of the input in which the clause-initial position may exhibit a stronger association with a particular type compared to the grammar that produced the input. That is, children would change the language as learners. As learners grow up, they themselves will become the source for the next generation and the cycle restarts.[3]

[1] An intriguing question related to this is whether a discrete threshold of variability exists after which the acquisition of V2 becomes impossible.

[2] I stay neutral as to the state of the grammatical representations in the mind of speakers. In principle, two fundamentally different options are available. It could be argued that the aforementioned changes only affect the E-language but not the I-language (Lightfoot 1999, 2006). This would entail that the grammar itself remains unaltered. Alternatively, the representations could change (Bybee 2006). In this case, the grammar would also change with the language usage by adults. Irrespective of the exact nature of the grammatical representations however, the shifts in the adult language will have an effect on the next generation of learners.

[3] Admittedly, adult L2/L*n* learners could also be considered as contributors to the change in the context of V2. As argued above, no differences are expected in the learning responses of children and adults. That is, adult learners should be equally susceptible to declining variability in clause-initial position. This is not so say that the acquisition process in children and adults will result in the same outcome. Plenty of evidence has been adduced that the grammars of L1 and L2/L*n* learners diverge (e.g.

5.2.2 Methodological consequences

Although my main contribution in this thesis is to the debate on the loss of V2, I also hope to provide valuable insights for the methodological advancement of ALL. As discussed in §1.6, most studies investigating syntactic phenomena have focused on patterns that are potentially less intricate in their derivation than V2 (e.g. basic word order and the word order in the nominal phrase). The experiments reported here as well as the work by Rebuschat (2008) and colleagues, and Getz (2018) demonstrate that ALL is indeed suitable for investigating syntactically more complex phenomena: Participants were able to learn an (semi)artificial V2 language. There are, however, certain caveats that need to be taken into account. First, participants' native language could be more prone to interfere with such phenomena. While Culbertson & Adger (2014) and Martin et al. (2020) found the same word order preferences in the noun phrase irrespective of the nature of the language (i.e. semi-artificial or artificial),[4] the set of current studies suggests that a significant influence of learners' L1 is indeed possible when semi-artificial languages are used. This became evident in the comparison of experiments 3 and 5.

A second caveat can be found in the size of the lexicon. One of the suggested reasons for the null result in experiment 4 was the small number of lexical items. That is, the variability did not suffice to make participants abandon their native SVO grammar. Even though this could possibly be an issue specific to V2 and the experiment design employed here, it still calls for a cautious approach. Experimenters need to consider beforehand whether the size of the lexicon is large enough and whether additional cues can be incorporated. The latter option may not always be possible (as in the case of this thesis). Furthermore, there are practical limits as to what participants can learn in one experimental sitting. Multi-day experiments have been proven to be possible (Hudson Kam & Newport 2005, 2009), but they might not always be feasible, for instance due to financial restrictions or due to unwanted sleep consolidation that occurs between experimental sessions (Gómez, Bootzin & Nadel 2006, Hupbach et al. 2009, Kim & Fenn 2020).

To mitigate these issues in the present thesis, I adopted iconic ALL in experiment 5 (§4.4). On the one hand, the work by Shapiro & Steinert-Threlkeld (2023) suggests that this paradigm reduces the influence of participant's L1. Their observation has received additional support by the findings of experiment 5. On the other hand, iconic ALL enables experimenters to use large lexicons without any prior training. Participants can focus on the relevant syntactic structures from the start. Given these advantages, any future investigations into complex syntactic pattern should therefore at least contemplate adopting iconic ALL. There are of course many open questions with respect to this novel paradigm and additional work is necessary. For instance, exclusively English speakers have been tested so far. Speakers of other languages might behave differently. It is also far from clear what kind of representations participants form during learning. Are they

Clahsen & Muysken 1986, Bohnacker & Rosén 2008, Bohnacker 2010). Nonetheless, it might be more appropriate to distinguish between learners and users instead of adults and children.

[4] See also Martin et al. (2024) for a replication with speakers of Kîîtharaka which is characterised by a non-homomorphic word order in the noun phrase.

completely on a par with language (spoken and signed) or do they feature icon-specific idiosyncrasies? Moreover, some fine-tuning, at least for the design in experiment 5, is required considering the observation of task-related effects. The results of Shapiro & Steinert-Threlkeld (2023) and the current thesis are nevertheless promising and encourage future use of iconic ALL.

5.3 Future work

The empirical studies presented in this thesis provide compelling evidence that learners of V2 languages are susceptible the variability in the clause-initial position and that too low variability can lead to the loss of V2. There are, however, some potentially relevant factors whose effects on learning should be scrutinised more closely. In what follows, I will discuss various options for future work, both experiment-based (§5.3.1) and corpus-based (§5.3.2).

5.3.1 Experimental work

Language is an intricate system with complex, interacting processes. Hence, aspects that play a role in natural languages need to be broken down and simplified or even completely omitted to successfully conduct experiments. In the context of the present set of studies, information-structural factors have been left aside. Information structure has, however, been shown to influence the placement of constituents in the clause-initial position. This is true for historic V2 languages (van Kemenade & Westergaard 2012, Steiner 2014, Galves 2020) but also for modern V2 languages (Frey 2006a, 2010, Light 2012). In Swedish, for instance, objects realised in clause-initial position may not be focused (Holmberg 2015: 348). Future experiments should thus test how learning will be affected when the clause-initial constituents are not placed 'randomly' in clause-initial position but when their placement is driven by their information-structural properties. I do not expect that significant differences to the effects observed here should arise. Learners still need to deduce from the input that no constraints apply to the fronting of constituents to the clause-initial position, even if fronting is information-structurally conditioned.

The experiments in Chapter 3 and Chapter 4 incorporated conditions where the distributions of clause-initial elements were skewed. The skew that was used in the experiments was somewhat unnatural (at least from a Germanic V2 perspective) in that the two non-dominant constituent types occurred with the same frequency in the clause-initial position. As the corpus work has revealed though, subjects and adjuncts are placed with considerable frequency in initial position (approximately 50–60% and 30–40% respectively), whereas all other constituent types occur much more infrequently (cf. §3.2). By adopting distributions observed for natural languages in the skewed conditions of experiments, a more natural skew could be used.[5] The acquisition of the language

[5]Some elements might be too infrequent to realistically feature at least once in the relatively small input learners receive. That is, some deviations from the skew observed in natural V2 languages might still be necessary. Besides, not all types of constituents appearing in initial position can be taken

could then be compared to the acquisition of a language with a skew attested in the languages that were in the process of losing V2. This is relevant for two reasons: People critical of ALL might object that the experiments are not ecological valid due to the unnatural skews. More importantly however, if the language with a skew comparable to the Modern Germanic language was still learnable but not the other language, further evidence in favour of my hypothesis would be obtained.

Another version of the experiment could also incorporate different types of adjuncts. Many of the adjuncts used in experiments 3–5 (e.g *in Boston*, *in the city*) could be interpreted as scene-setters (p.c. George Walkden). The discussion of experiment 5 (§4.4.4) has highlighted the cross-linguistically exceptional position of clause-initial scene-setters. The use of scene-setters both in training and testing might thus have affected participants' performance. In a novel iteration of the experiment, scene-setters could be replaced by other types of adjuncts that cannot be interpreted as scene-setters.

Another aspect that has not been considered in the experiments is the role of the frequency of particular lexical items. It has been a well-known fact since Zipf (1935) that lexical items are not uniformly distributed in natural languages. While few elements are used very frequently in natural languages, the remaining elements occur very infrequently. Such distributions are not confined to adult-directed speech but have also been noted for child-directed speech in several languages (Lavi-Rotbain & Arnon 2023). Crucially, Zipfian distributions have been shown to benefit learning in different contexts such as word segmentation (Kurumada, Meylan & Frank 2013, Lavi-Rotbain & Arnon 2022) and cross-situational learning of word meanings (Hendrickson & Perfors 2019).[6] Similar benefits on learning have also been noted for other skewed distributions that are not Zipfian (Wonnacott, Brown & Nation 2017). Interestingly, the frequency of lexical items can also affect how well syntactic constructions are learnt. Adults and children learn verb constructions better when the verbs in these constructions are skewed (Goldberg, Casenhiser & Sethuraman 2004, Casenhiser & Goldberg 2005). This poses the question whether skew in the lexical items occurring in initial position in V2 clauses can affect the acquisition of a V2 grammar. At least in Danish, the different lexical items that are realised in the clause-initial position are heavily skewed. According to Puggaard (2019: 299), the three most frequent elements are *det* 'it, that' (27.8%), *så* 'then' (18.6%) and *jeg* 'I' (15.2%). The training materials in experiments 3 and 5 on the other hand were perfectly uniform. That is, all lexical items occurred with the same frequency. It would thus be interesting to study how a skew in the lexical items would affect the learning performance (if at all). If differences were observed, such a skew could also function as further evidence for V2.

Apart from these relatively minor changes, more significant modifications to the experiment design could be made. Recall from §1.5.2 that different structures might function as evidence for a V2 grammar, such as alternating verb positions (V2/V-final) or particle verb constructions. Even though variability in the clause-initial position does

into consideration. For instance, non-finite verbs might be difficult to include in an iconic artificial language as they cannot easily be distinguished from finite verbs based on their icons alone.

[6]Interestingly, the learning advantage of such distributions might be domain-general as visual statistical learning appears to benefit from Zipfian distributions as well (Lavi-Rotbain & Arnon 2021).

affect the acquisition of a V2 grammar, contributions from other types of evidence were never ruled out. In fact, redundancy is a characteristic feature of languages and can benefit their acquisition (Tal & Arnon 2022). The introduction of additional cues might therefore lead to better learning. Alternatively, additional evidence could make learners less reliable on variability in the initial position. Related to this option would be the introduction of counter evidence to V2. For several of the languages discussed in §1.3, not only a rise in subject-initial sentences but also a rise in V3 sentences was noted. Such sentences exhibit a crucial role for Yang (2000, 2002) and others in their accounts for the loss of V2. It would thus be important to examine what effect different amounts of counter evidence would have on the acquisition of the languages. This option could also be further extended to test the effect of sociolinguistic aspects, such as language contact (Kroch, Taylor & Ringe 2000). The learning input could be provided by two competing social groups, similar to the design of Sneller & Roberts (2018). One of these groups would use a dialect with a strict V2 grammar, whereas the other group uses a dialect with a relaxed V2 grammar.

At the population level at least, V2 is not lost suddenly. This fact is reflected in the gradual decline of V2 structures noted in §1.3. That is, the loss of V2 was a process that spanned multiple generations. The experiments presented in Chapters 3 and 4 simulated the effect of variability in a single generation. Although learning was negatively affected when the variability was low, it does not become clear how the language would develop when transmission occurs from one generation to the next. Although one would expect a trajectory resembling the ones observed in natural languages, there is no guarantee to actually find such a trajectory. Besides, cultural transmission has been recognised as force that can shape language (Kirby, Cornish & Smith 2008, Kirby et al. 2015, Smith & Wonnacott 2010, Beckner, Pierrehumbert & Hay 2017, Saldana et al. 2019). Hence, a lab-based replication of such diachronic developments of natural languages would provide stronger evidence for the hypothesis under investigation in this thesis. The iterated learning paradigm developed by Kirby, Cornish & Smith (2008) would constitute a suitable methodology for this purpose: the output of one generation of participants is used as input for the next generation of participants. That is, the trajectory of the loss of V2 should become traceable with this paradigm. A potentially interesting version would be to compare trajectories of languages with and without additional evidence for V2. For instance, one group would be exposed to a variability in the clause-initial position as the only source of evidence. Another group would be exposed to variability and additionally to alternating verb positions (i.e. V2 and V-final).[7]

All of the sketched alternatives could also be conducted with different populations. On the one hand, children could be used as participants. On the other hand, different adult populations should be studied. The difficulty of acquiring languages in which objects account for the most frequent clause-initial constituent type could still be driven by a bias against initial objects (cf. Martin et al. (2024) for a similar argument for the

[7] Among the historical languages that have lost V2, English is the only language with a VO/OV alternation. At the time of the loss of V2, the loss of OV was already underway and relatively far advanced (Wallenberg et al. 2021: 6). It therefore seems reasonable to predict that the language with the both cues is more likely to sustain a V2 grammar.

word order in nominal phrases). Testing speakers of a language with OSV or OVS word order should not exhibit the same bias. Alternatively, languages with free word order could be used if objects occur with considerable frequency in initial position in these languages.

5.3.2 Corpus work

Apart from the aforementioned experimental work, additional corpus work is also desirable. If variability is crucial for the acquisition of V2, as the present thesis indeed suggests, all V2 languages should feature a relatively high proportion of adjunct-initial sentences. The discussion of previous corpus work in §3.2 has shown that this prediction is likely borne out, at least for some languages (e.g. Bohnacker & Rosén 2008, Bohnacker & Lindgren 2014, Puggaard 2019). However, those studies are based on a relatively small sample and should thus be cautiously interpreted. The validity of these findings could be probed with large-scale corpus studies, similar to the one I reported in §3.2 for German. Moreover, the distribution of clause-initial constituents has not been investigated for V2 languages of non-Germanic origin. Data from these languages would be particularly important as the fostering effect of variability should hold cross-linguistically irrespective of language (sub)families. The Universals Dependencies (UD) project provides treebanks from a host of different languages and might therefore constitute a suitable starting point. As of December 2023, the most recent UD release comprises the following V2 languages: Afrikaans, Breton, Danish, Dutch, Faroese, German, Icelandic, Norwegian, Swedish and Swiss German. Although the listed languages are mostly of Germanic origin, insights into the distribution in Breton would already extend existing knowledge considerably. Besides, the distributions of clause-initial constituents have not been investigated before for all of the V2 languages contained in UD. One potential caveat can found in the size of some these corpora in that they may contain only a small number of sentences (e.g. 100 sentences in the case of Swiss German). Nonetheless, any additional data would advance existing knowledge.

Even though variability in the clause-initial position fosters learning, there is no guarantee that this variability is actually present in the input child learners receive. That is, it is conceivable that V2 is acquired despite the absence of sufficient variability in the clause-initial position. To date, corpus-based studies mainly scrutinised adult-directed speech (though see Westergaard (2009b) for an exception). Although no obvious reasons can be conceived as to why the patterns observed in adult-directed speech would not hold for child-directed speech, the possibility of such differences can also not be immediately ruled out. A comparison of previous work but also my work (Table 3.4 and Figure 3.8 in §3.2) reveals that significant differences exists within languages. For example, the proportion of subject-initial sentences ranges from 50% to almost 67% in German, depending on the respective corpus. Accordingly, the proportion of adjunct-initial sentences and to some extent the proportion of object-initial sentences varies as well. That is, the distribution of clause-initial constituents appears to be context-sensitive. In this light, different distributions of clause-initial constituents might prevail in child-directed speech. This concern can be alleviated by examining the V2

child-direct speech in the CHILDES corpus (MacWhinney 2000). The presence of a considerable proportion of adjuncts would provide additional support for the role of variability.

In addition, the distributions of clause-initial elements in child-direct speech for individual children could be compared with the V2 errors children make during acquisition. A correlation between the two factors would provide further support for the role of variability in the input. This approach would be reminiscent of the study of Westergaard, Lohndal & Lundquist (2023) who compare the amount of non-subject-initial V2 sentences with the amount of V2 errors in the productions of Norwegian heritage language speakers.

5.4 Conclusion

My goal in this thesis is to provide evidence that general features of human cognition active during learning can (at least partially) explain the loss of V2 in different languages. Based on a domain-general fostering effect of variability on learning, I hypothesised that variability in the clause-initial constituent of V2 clauses fosters learning. Consequently, a lack of variability should lead to the loss of V2. The hypothesis was addressed in a series of artificial language learning experiments. As predicted, learners of a language with less variability in the clause-initial position showed a worse learning outcome than learners of a language with a higher variability in initial position. I interpret these results as evidence that a decline in variability was one of the drivers of the loss of V2. What is more, the results highlight that historical linguists and experimentally-working linguists could fruitfully collaborate more closely in the future. Evidence from both areas can help us to gain a better understanding of factors that underlie language change.

BIBLIOGRAPHY

Adams, Marianne. 1987a. From Old French to the theory of pro-drop. *Natural Language and Linguistic Theory* 5(1). 1–32. doi:10.1007/BF00161866.

Adams, Marianne. 1987b. *Old French, null subjects, and verb second phenomena*. Los Angeles: University of California dissertation.

Adelman, James S., Gordon D.A. Brown & José F. Quesada. 2006. Contextual diversity, not word frequency, determines word-naming and lexical decision times. *Psychological Science* 17(9). 814–823. doi:10.1111/j.1467-9280.2006.01787.x.

Andersen, Torben. 2017. Clause-internal antitopics in Berta, a Nilo-Saharan verb-second language. *Studies in Language* 41(1). 99–141. doi:10.1075/sl.41.1.04and.

Anderssen, Merete & Kristine Bentzen. 2018. Different outcomes in the acquisition of residual V2 and do-support in three Norwegian-English bilinguals: Cross-linguistic influence, dominance and structural ambiguity. *Frontiers in Psychology* 9. 1–18. doi:10.3389/fpsyg.2018.02130.

Anderssen, Merete & Marit Westergaard. 2010. Frequency and economy in the acquisition of variable word order. *Lingua* 120(11). 2569–2588. doi:10.1016/j.lingua.2010.06.006.

de Andrade, Aroldo. 2018. Aboutness topics in Old and Middle French: A corpus-based study on the fate of V2. *Canadian Journal of Linguistics* 63(2). 194–220. doi:10.1017/cnj.2017.45.

de Andrade, Aroldo & Charlotte Galves. 2019. Contrast and word order: A case study on the history of Portuguese. *Glossa: a journal of general linguistics* 4(1). 1–29. doi:10.5334/gjgl.897.

Andrason, Alexander. 2020. Verb second in Wymysorys. In Rebecca Woods & Sam Wolfe (eds.), *Rethinking verb second*, 700–722. Oxford: Oxford University Press. doi:10.1093/oso/9780198844303.003.0030.

Arechar, Antonio A. & David G. Rand. 2021. Turking in the time of COVID. *Behavior Research Methods* 53(6). 2591–2595. doi:10.3758/s13428-021-01588-4.

Aslin, Richard N., Jenny R. Saffran & Elissa L. Newport. 1998. Computation of conditional probability statistics by 8-month-old infants. *Psychological Science* 9(4). 321–324. doi:10.1111/1467-9280.00063.

Atkinson, Mark, Simon Kirby & Kenny Smith. 2015. Speaker input variability does not explain why larger populations have simpler languages. *PLoS ONE* 10(6). 1–20. doi:10.1371/journal.pone.0129463.

Atkinson, Mark, Kenny Smith & Simon Kirby. 2018. Adult learning and language simplification. *Cognitive Science* 42(8). 2818–2854. doi:10.1111/cogs.12686.

Axel, Katrin. 2007. *Studies on Old High German syntax. Left sentence periphery, verb placement and verb-second*. Amsterdam, Philadelphia: John Benjamins. doi:10.1075/la.112.

Axel, Katrin. 2009a. The consolidation of verb-second in Old High German: What role did subject pronouns play? In Paola Crisma & Giuseppe Longobardi (eds.), *Historical syntax and linguistic theory*, 128–143. Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780199560547.003.0008.

Axel, Katrin. 2009b. The verb-second property in Old High German: Different ways of filling the prefield. In Roland Hinterhölzl & Svetlana Petrova (eds.), *Information structure and language change: New approaches to word order variation in Germanic*, 17–44. Berlin, New York: Mouton de Gruyter. doi:10.1515/9783110216110.1.17.

Axel-Tober, Katrin. 2015. Historische Syntax des Deutschen aus generativer Perspektive. *Zeitschrift für germanistische Linguistik* 43(3). 450–484. doi:10.1515/zgl-2015-0024.

Axel-Tober, Katrin. 2018. Origins of verb-second in Old High German. In Agnes Jäger, Gisella Ferraresi & Helmut Weiß (eds.), *Clause structure and word order in the history of German*, 22–47. Oxford: Oxford University Press. doi:10.1093/oso/9780198813545.003.0003.

Bader, Markus. 2020. Objects in the German prefield. In Rebecca Woods & Sam Wolfe (eds.), *Rethinking verb second*, 15–39. Oxford: Oxford University Press. doi:10.1093/oso/9780198844303.003.0002.

Bader, Markus & Yvonne Portele. 2021. Discourse and form constraints on licensing object-first sentences in German. *Languages* 6(2). 1–28. doi:10.3390/languages6020082.

Barcroft, Joe & Mitchell S. Sommers. 2005. Effects of acoustic variability on second language vocabulary learning. *Studies in Second Language Acquisition* 27(3). 387–414. doi:10.1017/S0272263105050175.

Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1–48. doi:10.18637/jss.v067.i01.

Bech, Kristin. 2001. *Word order patterns in Old and Middle English: A syntactic and pragmatic study*. Bergen: University of Bergen dissertation.

Bech, Kristin & Christine Meklenborg Salvesen. 2014. Preverbal word order in Old English and Old French. In Kristin Bech & Kristine Gunn Eide (eds.), *Information structure and syntactic change in Germanic and Romance languages*, 233–269. Amsterdam, Philadelphia: John Benjamins. doi:10.1075/la.213.09bec.

Beckner, Clay, Janet B. Pierrehumbert & Jennifer Hay. 2017. The emergence of linguistic structure in an online iterated learning task. *Journal of Language Evolution* 2(2). 160–176. doi:10.1093/jole/lzx001.

Bell, Philippa K. 2017. Explicit and implicit learning: Exploring their simultaneity and immediate effectiveness. *Applied Linguistics* 38(3). 297–317. doi:10.1093/applin/amv028.

Benincà, Paola. 1995. Complement clitics in medieval Romance: The Tobler-Mussafia law. In Adrian Battye & Ian Roberts (eds.), *Clause structure and language change*, 325–

344. New York, Oxford: Oxford University Press. doi:10.1093/oso/9780195086324.003.0012.
Benincà, Paola. 2006. A detailed map of the left periphery of Medieval Romance. In Raffaella Zanuttini, Héctor Campos, Elena Herburger & Paul H. Portner (eds.), *Crosslinguistic research in syntax and semantics. Negation, tense and clausal architecture*, 53–86. Washington, DC: Georgetown University Press.
Benincà, Paola & Cecilia Poletto. 2004. Topic, focus, and V2. Defining the CP sublayers. In Luigi Rizzi (ed.), *The structure of CP and IP: The cartography of syntactic structures*, 52–75. Oxford, New York: Oxford University Press. doi:10.1093/oso/9780195159486.003.0003.
Bergmann, Christina, Sho Tsuji, Page E. Piccinini, Molly L. Lewis, Mika Braginsky, Michael C. Frank & Alejandrina Cristia. 2018. Promoting replicability in developmental research through meta-analyses: Insights from language acquisition research. *Child Development* 89(6). 1996–2009. doi:10.1111/cdev.13079.
Berwick, Robert C. 1985. *The acquisition of syntactic knowledge*. Cambridge, MA, London: MIT Press. doi:10.7551/mitpress/1074.001.0001.
den Besten, Hans. 1983. On the interaction of root transformations and lexical deletive rules. In Werner Abraham (ed.), *On the formal syntax of the Westgermania*, 47–131. Amsterdam: John Benjamins. doi:10.1075/la.3.03bes.
den Besten, Hans. 1989. *Studies in West Germanic syntax*. Tilburg: Katholieke Universiteit Brabant Proefschrift.
Bhatt, Rakesh Mohan. 1999. *Verb movement and the syntax of Kashmiri*. Dordrecht, Boston, London: Kluwer. doi:10.1007/978-94-015-9279-6.
Biberauer, Theresa & Ans van Kemenade. 2011. Subject positions and information-structural diversification in the history of English. *Catalan Journal of Linguistics* 10(1). 17–69. doi:10.5565/rev/catjl.32.
Birner, Betty. 1995. Pragmatic constraints on the verb in English inversion. *Lingua* 97(4). 233–256. doi:10.1016/0024-3841(95)00026-7.
Bohnacker, Ute. 2010. The clause-initial position in L2 Swedish declaratives: Word order variation and discourse pragmatics. *Nordic Journal of Linguistics* 33(2). 105–143. doi:10.1017/S033258651000017X.
Bohnacker, Ute & Josefin Lindgren. 2014. Fundament, formellt subjekt och frekvens: Ordföljdsmönster i svenska, nederländska och hos vuxna inlärare av svenska. *Språk & Stil* 24. 33–71.
Bohnacker, Ute & Christina Rosén. 2008. The clause-initial position in L2 German declaratives. Transfer of information structure. *Studies in Second Language Acquisition* 30(4). 511–538. doi:10.1017/S0272263108080741.
Borsley, Robert D. & Andreas Kathol. 2000. Breton as a V2 language. *Linguistics* 38(4). 665–710. doi:10.1515/ling.2000.002.
Bošković, Željko. 2020. On the syntax and prosody of verb second and clitic second. In Rebecca Woods & Sam Wolfe (eds.), *Rethinking verb second*, 503–535. Oxford: Oxford University Press. doi:10.1093/oso/9780198844303.003.0021.
Bradlow, Ann R. & Tessa Bent. 2008. Perceptual adaptation to non-native speech. *Cognition* 106(2). 707–729. doi:10.1016/j.cognition.2007.04.005.

Braine, Martin D. S. 1963. On learning the grammatical order of words. *Psychological Review* 70(4). 323–348. doi:10.1037/h0047696.
Braine, Martin D. S. 1966. Learning the positions of words relative to a marker element. *Journal of Experimental Psychology* 72(4). 532–540. doi:10.1037/h0023763.
Branigan, Holly P., Martin J. Pickering & Mikihiro Tanaka. 2008. Contributions of animacy to grammatical function assignment and word order during production. *Lingua* 118(2). 172–189. doi:10.1016/j.lingua.2007.02.003.
Branigan, Philip. 1996. Verb-second and the A-bar syntax of subjects. *Studia Linguistica* 50(1). 50–79. doi:10.1111/j.1467-9582.1996.tb00343.x.
Bross, Fabian. 2020. Encoding different types of topics and foci in German Sign Language. A cartographic approach to sign language syntax. *Glossa: a journal of general linguistics* 5(1). 1–29. doi:10.5334/gjgl.1094.
Bruening, Benjamin. 2010. Language-particular syntactic rules and constraints: English locative inversion and do-support. *Language* 86(1). 43–84. doi:10.1353/lan.0.0201.
Bruening, Benjamin. 2016. Alignment in syntax: Quotative inversion in english. *Syntax* 19(2). 111–155. doi:10.1111/synt.12121.
Bybee, Joan. 2006. From usage to grammar: The mind's response to repetition. *Language* 82(4). 711–733. doi:10.1353/lan.2006.0186.
Bybee, Joan. 2009. Grammaticization: Implications for a theory of language. In Jiansheng Guo, Elena Lieven, Nancy Budwig, Susan Ervin-Tripp, Keiko Nakamura & Seyda Ozcaliskan (eds.), *Crosslinguistic approaches to the psychology of language: Research in the tradition of Dan Isaac Slobin*, 345–355. New York, London: Psychology Press. doi:10.4324/9780203837887.
Bybee, Joan L. & Dan I. Slobin. 1982. Rules and schemas in the development and use of the English past tense. *Language* 58(2). 265–289. doi:10.1353/lan.1982.0021.
Cardinaletti, Anna. 1990. Subject/object asymmetries in German null-topic constructions and the status of specCP. In Joan Mascaró & Marina Nespor (eds.), *Grammar in progress. Glow essays for Henk van Riemsdijk*, 75–84. Dordrecht, Providence RI: Foris. doi:10.1515/9783110867848.75.
Casalicchio, Jan & Federica Cognola. 2018. Verb-second and (micro)variation in two Rhaeto-Romance varieties of Northern Italy. In Roberta D'Alessandro & Diego Pescarini (eds.), *Advances in Italian dialectology. Sketches of Italo-Romance grammars*, 72–105. Leiden, Boston: Brill. doi:10.1163/9789004354395.
Casalicchio, Jan & Federica Cognola. 2020. Parameterizing subject–verb inversion across verb second languages: On the role of relativized minimality at the vP edge. In Rebecca Woods & Sam Wolfe (eds.), *Rethinking verb second*, 594–622. Oxford: Oxford University Press. doi:10.1093/oso/9780198844303.003.0025.
Casenhiser, Devin & Adele E. Goldberg. 2005. Fast mapping between a phrasal form and meaning. *Developmental Science* 8(6). 500–508. doi:10.1111/j.1467-7687.2005.00441.x.
Cavalcante, Silvia Regina de Oliveira, Charlotte Galves & Maria Clara Paixão de Sousa. 2015. Topics, subjects and grammatical change: From Classical to Modern European Portuguese. *LaborHistórico* 1(2). 97–107. doi:10.24206/lh.v1i2.4796.

Chater, Nick & Paul Vitányi. 2003. Simplicity: A unifying principle in cognitive science? *Trends in Cognitive Sciences* 7(1). 19–22. doi:10.1016/S1364-6613(02)00005-0.
Chomsky, Noam. 1986. *Knoweldge of language. It's nature, origin and use*. New York: Praeger.
Chomsky, Noam. 1993. *Lectures on government and binding: The Pisa lectures*. Berlin, New York: Mouton de Gruyter. doi:10.1515/9783110884166.
Chomsky, Noam. 2005. Three factors in language design. *Linguistic Inquiry* 36(1). 1–22. doi:10.1162/0024389052993655.
Cichosz, Anna. 2010. *Influence of text type on word order of old Germanic languages. A corpus-based contrastive study of Old English and Old High German*. Frankfurt am Main: Peter Lang.
Clahsen, Harald. 1990. The comparative study of first and second language development. *Studies in Second Language Acquisition* 12(2). 135–153. doi:10.1017/S0272263100009050.
Clahsen, Harald, Sonja Eisenbeiss & Martina Penke. 1996. Lexical learning in early syntactic development. In Harald Clahsen (ed.), *Generative perspectives on language acquisition. Empirical findings, theoretical considerations and crosslinguistic comparisons*, 129–159. Amsterdam, Philadelphia: John Benjamins. doi:10.1075/lald.14.08cla.
Clahsen, Harald & Pieter Muysken. 1986. The availability of Universal Grammar to adult and child learners – a study of the acquisition of German word order. *Second Language Research* 2(2). 93–119. doi:10.1177/026765838600200201.
Claidière, Nicolas, Gameli Kodjo-kuma Amedon, Jean Baptiste André, Simon Kirby, Kenny Smith, Dan Sperber & Joël Fagot. 2018. Convergent transformation and selection in cultural evolution. *Evolution and Human Behavior* 39(2). 191–202. doi:10.1016/j.evolhumbehav.2017.12.007.
Clark, Robin & Ian Roberts. 1993. A computational model of language learnability and language change. *Linguistic Inquiry* 24(2). 299–345.
Clopper, Cynthia G. & David B. Pisoni. 2004. Effects of talker variability on perceptual learning of dialects. *Language and Speech* 47(3). 207–238. doi:10.1177/00238309040470030101.
Cognola, Federica. 2013. The mixed OV/VO syntax of Mòcheno main clauses: On the interaction between high and low left periphery. In Theresa Biberauer & Michelle Sheehan (eds.), *Theoretical approaches to disharmonic word order*, 106–135. Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780199684359.003.0004.
Cognola, Federica. 2015. How to define relaxed V2 languages and how to distinguish them from non-V2 languages: A reply to Brandtler (2014). *Nordic Journal of Linguistics* 38(1). 93–100. doi:10.1017/S0332586515000037.
Cognola, Federica. 2019. On the left periphery of three languages of Northern Italy. *Linguistic Variation* 19(1). 82–117. doi:10.1075/lv.16005.cog.
Collins, Chris & Phil Branigan. 1997. Quotative inversion. *Natural Language and Linguistic Theory* 15(1). 1–41. doi:10.1023/A:1005722729974.
Cournane, Ailís. 2017. In defence of the child innovator. In Éric Mathieu & Robert Truswell (eds.), *Micro-change and macro-change in diachronic syntax*, 10–24. Oxford: Oxford University Press. doi:10.1093/oso/9780198747840.003.0002.

Cournane, Ailís. 2019. A developmental view on incrementation in language change. *Theoretical Linguistics* 45(3). 127–150. doi:10.1515/tl-2019-0010.

Cournane, Ailís & Espen Klævik-Pettersen. 2023. The role of the conservative learner in the rise and fall of verb-second. *Journal of Historical Syntax* 7(9). 1–48. doi:10.18148/hs/2023.v7i6-19.160.

Crump, Matthew J. C., John V. McDonnell & Todd M. Gureckis. 2013. Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. *PLoS ONE* 8(3). 1–18. doi:10.1371/journal.pone.0057410.

Cruschina, Silvio & Craig Sailor. 2022. What is "residual verb second"? And what does Romance have to do with it? *Isogloss. Open Journal of Romance Linguistics* 8(3). 1–28. doi:10.5565/rev/isogloss.235.

Culbertson, Jennifer. 2012. Typological universals as reflections of biased learning: Evidence from artificial language learning. *Language and Linguistics Compass* 6(5). 310–329. doi:10.1002/lnc3.338.

Culbertson, Jennifer. 2017. New approaches to Greenbergian word order dependencies. In N. J. Enfield (ed.), *Dependencies in language: On the causal ontology of linguistic systems*, 23–38. Berlin: Language Science Press. doi:10.5281/zenodo.573778.

Culbertson, Jennifer. 2023. Artificial language learning. In Jon Sprouse (ed.), *The Oxford handbook of experimental syntax*, 271–300. Oxford: Oxford University Press. doi:10.1093/oxfordhb/9780198797722.013.9.

Culbertson, Jennifer & David Adger. 2014. Language learners privilege structured meaning over surface frequency. *Proceedings of the National Academy of Sciences* 111(16). 5842–5847. doi:10.1073/pnas.1320525111.

Culbertson, Jennifer, Julie Franck, Guillaume Braquet, Magda Barrera Navarro & Inbal Arnon. 2020. A learning bias for word order harmony: Evidence from speakers of non-harmonic languages. *Cognition* 204. 1–18. doi:10.1016/j.cognition.2020.104392.

Culbertson, Jennifer & Elissa L. Newport. 2015. Harmonic biases in child learners: In support of language universals. *Cognition* 139. 71–82. doi:10.1016/j.cognition.2015.02.007.

Culbertson, Jennifer & Elissa L. Newport. 2017. Innovation of word order harmony across development. *Open Mind* 1(2). 91–100. doi:10.1162/opmi_a_00010.

Culbertson, Jennifer, Marieke Schouwstra & Simon Kirby. 2020. From the world to word order: Deriving biases in noun phrase order from statistical properties of the world. *Language* 96(3). 696–717. doi:10.1353/lan.2020.0045.

Culbertson, Jennifer & Kathryn Schuler. 2019. Artificial language learning in children. *Annual Review of Linguistics* 5. 353–373. doi:10.1146/annurev-linguistics-011718-012329.

Culbertson, Jennifer, Paul Smolensky & Géraldine Legendre. 2012. Learning biases predict a word order universal. *Cognition* 122(3). 306–329. doi:10.1016/j.cognition.2011.10.017.

Diesing, Molly. 1990. Verb movement and the subject position in Yiddish. *Natural Language and Linguistic Theory* 8(1). 41–79. doi:10.1007/BF00205531.

Diessel, Holger. 2011. Grammaticalization and language acquisition. In Bernd Heine & Heiko Narrog (eds.), *The Oxford handbook of grammaticalization*, 130–141. Oxford, New York: Oxford University Press. doi:10.1093/oxfordhb/9780199586783.013.0011.
Doherty, Monika. 2005. Topic-worthiness in German and English. *Linguistics* 43(1). 181–206. doi:10.1515/ling.2005.43.1.181.
Drach, Erich. 1937. *Grundgedanken der deutschen Satzlehre*. Frankfurt am Main: Diesterweg.
Ehala, Martin. 2006. The word order of Estonian: Implications to universal language. *Journal of Universal Language* 7(1). 49–89. doi:10.22425/jul.2006.7.1.49.
Elazar, Amit, Raquel G. Alhama, Louisa Bogaerts, Noam Siegelman, Cristina Baus & Ram Frost. 2022. When the "tabula" is anything but "rasa:" What determines performance in the auditory statistical learning task? *Cognitive Science* 46(2). 1–17. doi:10.1111/cogs.13102.
Endress, Ansgar D., Brian J. Scholl & Jacques Mehler. 2005. The role of salience in the extraction of algebraic rules. *Journal of Experimental Psychology: General* 134(3). 406–419. doi:10.1037/0096-3445.134.3.406.
Engel, Ulrich. 1974. Syntaktische Besonderheiten der deutschen Alltagssprache. In Hugo Moser (ed.), *Gesprochene Sprache. Jahrbuch 1972*, 199–228. Düsseldorf: Schwann.
Esper, Erwin Allen. 1925. A technique for the experimental investigation of associative interference in artificial linguistic material. *Language Monographs* 1. 5–47.
Ettlinger, Marc, Kara Morgan-Short, Mandy Faretta-Stutenberg & Patrick C. M. Wong. 2016. The relationship between artificial and second language learning. *Cognitive Science* 40(4). 822–847. doi:10.1111/cogs.12257.
Eythórsson, Thórhallur. 1995. *Verbal syntax in the early Germanic languages*. Ithaca, NY: Cornell University dissertation.
Faarlund, Jan Terje. 2004. *The syntax of Old Norse. With a survey of inflectional morphology and a complete bibliography*. Oxford, New York: Oxford University Press. doi:10.1093/acprof:oso/9780199235599.001.0001.
Fedzechkina, Maryia, T. Florian Jaeger & Elissa L. Newport. 2012. Language learners restructure their input to facilitate efficient communication. *Proceedings of the National Academy of Sciences* 109(44). 17897–17902. doi:10.1073/pnas.1215776109.
Fedzechkina, Maryia, Elissa L. Newport & T. Florian Jaeger. 2017. Balancing effort and information transmission during language acquisition: Evidence from word order and case marking. *Cognitive Science* 41(2). 416–446. doi:10.1111/cogs.12346.
Fedzechkina, Masha & T. Florian Jaeger. 2020. Production efficiency can cause grammatical change: Learners deviate from the input to better balance efficiency against robust message transmission. *Cognition* 196. 1–11. doi:10.1016/j.cognition.2019.104115.
Ferdinand, Vanessa, Simon Kirby & Kenny Smith. 2019. The cognitive roots of regularization in language. *Cognition* 184. 53–68. doi:10.1016/j.cognition.2018.12.002.
Finn, Amy S. & Carla L. Hudson Kam. 2008. The curse of knowledge: First language knowledge impairs adult learners' use of novel statistics for word segmentation. *Cognition* 108(2). 477–499. doi:10.1016/j.cognition.2008.04.002.

Fischer, Olga, Ans van Kemenade, Willem Koopman & Wim van der Wurff. 2001. *The syntax of Early English*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511612312.

Fitch, W. Tecumseh & Marc D. Hauser. 2004. Computational constraints on syntactic processing in a nonhuman primate. *Science* 303(5656). 377–380. doi:10.1126/science.1089401.

Fitzgerald, Colleen M. 2020. Second and first position in Tohono O'odham auxiliaries. In Rebecca Woods & Sam Wolfe (eds.), *Rethinking verb second*, 745–769. Oxford: Oxford University Press. doi:10.1093/oso/9780198844303.003.0032.

Fodor, Janet Dean. 1998. Unambiguous triggers. *Linguistic Inquiry* 29(1). 1–36. doi:10.1162/002438998553644.

Folia, Vasiliki, Julia Uddén, Meinou De Vries, Christian Forkstam & Karl Magnus Petersson. 2010. Artificial language learning in adults and children. *Language Learning* 60. 188–220. doi:10.1111/j.1467-9922.2010.00606.x.

Fontana, Josep M. 1993. *Phrase structure and the syntax of clitics in the history of Spanish*. Philadelphia, PA: University of Pennyslvania Dissertation.

Fontana, Josep M. 1997. On the integration of second position phenomena. In Ans van Kemenade & Nigel Vincent (eds.), *Parameters of morphosyntactic change*, 207–249. Cambridge, New York, Melbourne: Cambridge University Press.

Frances, Candice, Clara D. Martin & Jon Andoni Duñabeitia. 2020. The effects of contextual diversity on incidental vocabulary learning in the native and a foreign language. *Scientific Reports* 10. 1–11. doi:10.1038/s41598-020-70922-1.

Franck, Julie, Irene Rotondi & Ulrich H. Frauenfelder. 2016. Learning structure-dependent agreement in a hierarchical artificial grammar. *Journal of Memory and Language* 87. 84–104. doi:10.1016/j.jml.2015.11.003.

Frascarelli, Mara & Roland Hinterhölzl. 2007. Types of topics in German and Italian. In Susanne Winkler & Kerstin Schwabe (eds.), *On information structure, meaning, and form*, 87–116. Amsterdam, Philadelphia: John Benjamins. doi:10.1075/la.100.07fra.

Freitag, Constantin. 2021. *Verb-second as a reconstruction phenomenon: Evidence from grammar and processing*. Berlin, Boston: De Gruyter. doi:10.1515/9783110725018.

Frey, Werner. 2006a. Contrast and movement to the German prefield. In Valéria Molnár & Susanne Winkler (eds.), *The architecture of focus*, 235–264. Berlin, New York: Mouton de Gruyter. doi:10.1515/9783110922011.235.

Frey, Werner. 2006b. How to get an object-es into the German prefield. In Patrick Brandt & Eric Fuß (eds.), *Form, structure, and grammar. A festschrift presented to Günther Grewendorf on occasion of his 60th birthday*, 159–186. Berlin: Akademie Verlag. doi:10.1524/9783050085555.159.

Frey, Werner. 2010. Ā-movement and conventional implicatures: About the grammatical encoding of emphasis in German. *Lingua* 120(6). 1416–1435. doi:10.1016/j.lingua.2008.09.016.

Freywald, Ulrike. 2008. Zur Syntax und Funktion von dass-Sätzen mit Verbzweitstellung. *Deutsche Sprache* 36(3). 246–285. doi:10.37307/j.1868-775X.2008.03.04.

Freywald, Ulrike. 2009. Kontexte für nicht-kanonische Verbzweitstellung. In Veronika Ehrich, Christian Fortmann, Ingo Reich & Marga Reis (eds.), *Koordination und Subordination im Deutschen*, 113–134. Hamburg: Buske.
Freywald, Ulrike. 2013. Uneingeleiteter V1- und V2-Satz. In Jörg Meibauer, Markus Steinbach & Hans Altmann (eds.), *Satztypen des Deutschen*, 317–337. Berlin, Boston: De Gruyter. doi:10.1515/9783110224832.317.
Freywald, Ulrike. 2016. „V2-Nebensätze“ – ein eigener Satztyp? In Rita Finkbeiner & Jörg Meibauer (eds.), *Satztypen und Konstruktionen*, 326–372. Berlin, Boston: De Gruyter. doi:10.1515/9783110423112-011.
Freywald, Ulrike, Leonie Cornips, Natalia Ganuza, Ingvild Nistov & Toril Opsahl. 2015. Beyond verb second – a matter of novel information-structural effects? Evidence from Norwegian, Swedish, German and Dutch. In Jacomine Nortier & Bente A. Svendsen (eds.), *Language, youth and identity in the 21st century*, 73–92. Camridge: Cambridge University Press. doi:10.1017/CBO9781139061896.006.
Friederici, Angela D., Jörg Bahlmann, Stefan Heim, Ricarda I. Schubotz & Alfred Anwander. 2006. The brain differentiates human and non-human grammars: Functional localization and structural connectivity. *Proceedings of the National Academy of Sciences* 103(7). 2458–2463. doi:10.1073/pnas.0509389103.
Friederici, Angela D., Karsten Steinhauer & Erdmut Pfeifer. 2002. Brain signatures of artificial language processing: Evidence challenging the critical period hypothesis. *Proceedings of the National Academy of Sciences* 99(1). 529–534. doi:10.1073/pnas.012611199.
Fuss, Eric. 2003. On the historical core of V2 in Germanic. *Nordic Journal of Linguistics* 26(2). 195–231. doi:10.1017/S0332586503001082.
Fuss, Eric. 2008. *Word order and language change. On the interface between syntax and morphology*. Frankfurt am Main: Johann-Wolfang-Goethe Universität zu Frankfurt am Main Habilitationsschrift.
Galle, Marcus E., Keith S. Apfelbaum & Bob McMurray. 2015. The role of single talker acoustic variation in early word learning. *Language Learning and Development* 11(1). 66–79. doi:10.1080/15475441.2014.895249.
Galves, Charlotte. 2018. The Tycho Brahe Corpus of Historical Portuguese: Methodology and results. *Linguistic Variation* 18(1). 49–73. doi:10.1075/lv.00004.gal.
Galves, Charlotte. 2020. Relaxed verb second in Classical Portuguese. In Rebecca Woods & Sam Wolfe (eds.), *Rethinking verb second*, 368–395. Oxford: Oxford University Press. doi:10.1093/oso/9780198844303.003.0016.
Galves, Charlotte, Helena Brito & Maria Clara Paixão de Sousa. 2005. The change in clitic placement from Classical to Modern European Portuguese: Results from the Tycho Brahe corpus. *Journal of Portuguese Linguistics* 4(2). 39–67. doi:10.5334/jpl.166.
Galves, Charlotte & Pablo Faria. 2017. *Tycho Brahe parsed corpus of Historical Portuguese*.
Galves, Charlotte & Alba Gibrail. 2018. Subject inversion in transitive sentences from Classical to Modern European Portuguese: A corpus-based study. In Ana Maria Martins & Adriana Cardoso (eds.), *Word order change*, 163–178. Oxford: Oxford University Press. doi:10.1093/oso/9780198747307.003.0009.

Galves, Charlotte & Anthony Kroch. 2016. Main syntactic changes from a principle-and-parameters view. In W. Leo Wetzels, João Costa & Sergio Menuzzi (eds.), *The handbook of Portuguese linguistics*, 487–503. Chichester: Wiley. doi:10.1002/9781118791844.ch27.

Galves, Charlotte & Maria Clara Paixão de Sousa. 2017. The change in the position of the verb in the history of Portuguese: Subject realization, clitic placement, and prosody. *Language* 93(3). e152–e180. doi:10.1353/lan.2017.0049.

Gao, Jianwu & Shuang Ma. 2021. Learning condition, linguistic complexity, and first language transfer in semiartificial language learning. A conceptual replication and extension of Tagarelli et al. (2016). *Studies in Second Language Acquisition* 43(2). 355–378. doi:10.1017/S0272263120000686.

van Gelderen, Elly. 2011. *The linguistic cycle: Language change and the language faculty*. Oxford, New York: Oxford University Press. doi:10.1093/acprof:oso/9780199756056.001.0001.

Gentner, Timothy Q., Kimberly M. Fenn, Daniel Margoliash & Howard C. Nusbaum. 2006. Recursive syntactic pattern learning by songbirds. *Nature* 440(7088). 1204–1207. doi:10.1038/nature04675.

Getz, Heidi. 2018. *Sentence first, arguments after: Mechanisms of morphosyntax acquisition*. Washington, DC: Georgetown University dissertation. doi:10.31237/osf.io/ske5q.

Getz, Heidi. 2019. Acquisition of morphosyntax: A pattern learning approach. In David W. Lightfoot & Jonathan Havenhill (eds.), *Variable properties in language. Their nature and acquisition*, 103–114. Washington, DC: Georgetown University Press.

Goldberg, Adele E., Devin M. Casenhiser & Nitya Sethuraman. 2004. Learning argument structure generalizations. *Cognitive Linguistics* 15(3). 289–316. doi:10.1515/cogl.2004.011.

Goldin-Meadow, Susan, Wing Chee So, Aslı Özyürek & Carolyn Mylander. 2008. The natural order of events: How speakers of different languages represent events nonverbally. *Proceedings of the National Academy of Sciences* 105(27). 9163–9168. doi:10.1073/pnas.0710060105.

Gómez, Rebecca & Jessica Maye. 2005. The developmental trajectory of nonadjacent dependency learning. *Infancy* 7(2). 183–206. doi:10.1207/s15327078in0702_4.

Gómez, Rebecca L. 2002. Variability and detection of invariant structure. *Psychological Science* 13(5). 431–436. doi:10.1111/1467-9280.00476.

Gómez, Rebecca L., Richard R. Bootzin & Lynn Nadel. 2006. Naps promote abstraction in language-learning infants. *Psychological Science* 17(8). 670–674. doi:10.1111/j.1467-9280.2006.01764.x.

Gómez, Rebecca L. & LouAnn Gerken. 2000. Infant artificial language learning and language acquisition. *Trends in Cognitive Sciences* 4(5). 178–186. doi:10.1016/S1364-6613(00)01467-4.

Graf Estes, Katharine & Casey Lew-Williams. 2015. Listening through voices: Infant statistical word segmentation across multiple speakers. *Developmental Psychology* 51(11). 1517–1528. doi:10.1037/a0039725.

Greco, Ciro & Liliane Haegeman. 2020. Frame setters and microvariation of subject-initial verb second. In Rebecca Woods & Sam Wolfe (eds.), *Rethinking verb second*, 61–89. Oxford: Oxford University Press. doi:10.1093/oso/9780198844303.003.0004.

Green, Georgia M. 1980. Some wherefores of English inversions. *Language* 56(3). 582–601. doi:10.2307/414451.
Grünert, Matthias. 2018. Die Konstituentenfolge im bündnerromanischen Nebensatz. *Ladinia* 42. 23–39.
Grunow, Hope, Tammie J. Spaulding, Rebecca L. Gómez & Elena Plante. 2006. The effects of variation on learning word order rules by adults with and without language-based learning disabilities. *Journal of Communication Disorders* 39(2). 158–170. doi:10.1016/j.jcomdis.2005.11.004.
de Haan, Germen. 2001. More is going on upstairs than downstairs: Embedded root phenomena in West Frisian. *Journal of Comparative Germanic Linguistics* 4(1). 3–38. doi:10.1023/A:1012224020604.
Hack, Franziska Maria & Georg A Kaiser. 2013. Zur Syntax von Fragesätzen im Rätoromanischen. In Georges Darms, Clà Riatsch & Clau Solèr (eds.), *Akten des V. rätoromanischen Kolloquiums/Actas dal V. Colloqui retoromanistic Lavin 2011*, 137–161. Tübingen: Francke.
Haeberli, Eric. 1999. On the word order 'XP-subject' in the Germanic languages. *Journal of Comparative Germanic Linguistics* 3(1). 1–36. doi:10.1023/A:1009889526118.
Haeberli, Eric. 2002a. Inflectional morphology and the loss of verb-second in English. In David W. Lightfoot (ed.), *Syntactic effects of morphological change*, 88–106. Oxford, New York: Oxford University Press. doi:10.1093/acprof:oso/9780199250691.003.0005.
Haeberli, Eric. 2002b. Observations on the loss of verb second in the history of English. In C. Jan-Wouter Zwart & Werner Abraham (eds.), *Studies in comparative Germanic syntax*, 245–272. Amsterdam, Philadelphia: John Benjamins. doi:10.1075/la.53.15hae.
Haeberli, Eric & Tabea Ihsane. 2016. Revisiting the loss of verb movement in the history of English. *Natural Language and Linguistic Theory* 34(2). 497–542. doi:10.1007/s11049-015-9312-x.
Haeberli, Eric, Susan Pintzuk & Ann Taylor. 2020. Object pronoun fronting and the nature of verb second in early English. In Rebecca Woods & Sam Wolfe (eds.), *Rethinking verb second*, 396–425. Oxford: Oxford University Press. doi:10.1093/oso/9780198844303.003.0017.
Hamrick, Phillip & Rebecca Sachs. 2018. Establishing evidence of learning in experiments employing artificial linguistics systems. *Studies in Second Language Acquisition* 40(1). 153–169. doi:10.1017/S0272263116000474.
Hendrickson, Andrew T. & Andrew Perfors. 2019. Cross-situational learning in a Zipfian environment. *Cognition* 189. 11–22. doi:10.1016/j.cognition.2019.03.005.
Hendriksen, Hans. 1990. Sentence position of the verb in Himachali. *Acta Linguistica Hafniensia* 22(1). 159–171. doi:10.1080/03740463.1990.10411526.
Henrich, Joseph, Steven J. Heine & Ara Norenzayan. 2010. The weirdest people in the world? *Behavioral and Brain Sciences* 33(2). 61–83. doi:10.1017/S0140525X0999152X.
Heycock, Caroline, Antonella Sorace & Zakaris Svabo Hansen. 2010. V-to-I and V2 in subordinate clauses: An investigation of Faroese in relation to Icelandic and Danish. *Journal of Comparative Germanic Linguistics* 13(1). 61–97. doi:10.1007/s10828-010-9035-7.

Heycock, Caroline & Joel Wallenberg. 2013. How variational acquisition drives syntactic change. *Journal of Comparative Germanic Linguistics* 16(2). 127–157. doi:10.1007/s10828-013-9056-0.

Hinterhölzl, Roland. 2009. The role of information structure in word order variation and word order change. In Roland Hinterhölzl & Svetlana Petrova (eds.), *Information structure and language change: New approaches to word order variation in Germanic*, 45–66. Berlin, New York: Mouton de Gruyter. doi:10.1515/9783110216110.1.45.

Hinterhölzl, Roland & Ans van Kemenade. 2012. The interaction between syntax, information structure and prosody in word order change. In Terttu Nevalainen & Elizabeth Closs Traugott (eds.), *The Oxford handbook of the history of English*, 803–821. New York: Oxford University Press. doi:10.1093/oxfordhb/9780199922765.013.0066.

Holmberg, Anders. 2015. Verb second. In Tibor Kiss & Artemis Alexiadou (eds.), *Syntax – Theory and analysis. An international handbook*, 342–383. Berlin, München, Boston: De Gruyter Mouton. doi:10.1515/9783110377408.342.

Holmberg, Anders & Christer Platzack. 1995. *The role of inflection in Scandinavian syntax*. New York, Oxford: Oxford University Press. doi:10.1093/oso/9780195067453.001.0001.

Holmberg, Anders, Heete Sahkai & Anne Tamm. 2020. Prosody distinguishes Estonian V2 from Finnish and Swedish. *Speech Prosody* 10. 439–443. doi:10.21437/SpeechProsody.2020-90.

Hopman, Elise W.M. & Maryellen C. MacDonald. 2018. Production practice during language learning improves comprehension. *Psychological Science* 29(6). 961–971. doi:10.1177/0956797618754486.

Horst, Jessica S. & Michael C. Hout. 2016. The Novel Object and Unusual Name (NOUN) Database: A collection of novel images for use in experimental research. *Behavior Research Methods* 48(4). 1393–1409. doi:10.3758/s13428-015-0647-3.

Hróarsdóttir, Thorbjörg. 2003. Language change and language acquisition. *Nordlyd* 31(1). 116–131. doi:10.7557/12.21.

Hsiao, Yaling & Kate Nation. 2018. Semantic diversity, frequency and the development of lexical quality in children's word reading. *Journal of Memory and Language* 103. 114–126. doi:10.1016/j.jml.2018.08.005.

Hudson Kam, Carla L. & Elissa L. Newport. 2005. Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language Learning and Development* 1(2). 151–195. doi:10.1080/15475441.2005.9684215.

Hudson Kam, Carla L. & Elissa L. Newport. 2009. Getting it right by getting it wrong: When learners change languages. *Cognitive Psychology* 59(1). 30–66. doi:10.1016/j.cogpsych.2009.01.001.

Hulk, Aafke & Ans van Kemenade. 1995. Verb second, pro-drop, functional projections and language change. In Adrian Battye & Ian Roberts (eds.), *Clause structure and language change*, 227–256. New York, Oxford: Oxford University Press. doi:10.1093/oso/9780195086324.003.0008.

Hulk, Aafke & Jean-Yves Pollock. 2001. Subject positions in Romance and the theory of Universal Grammar. In Aafke Hulk & Jean-Yves Pollock (eds.), *Subject inversion in*

Romance and the theory of Universal Grammar, 3–19. New York: Oxford University Press. doi:10.1093/oso/9780195142693.003.0001.
Hupbach, Almut, Rebecca L. Gómez, Richard R. Bootzin & Lynn Nadel. 2009. Nap-dependent learning in infants. *Developmental Science* 12(6). 1007–1012. doi:10.1111/j.1467-7687.2009.00837.x.
Hupp, Julie M., Vladimir M. Sloutsky & Peter W. Culicover. 2009. Evidence for a domain-general mechanism underlying the suffixation preference in language. *Language and Cognitive Processes* 24(6). 876–909. doi:10.1080/01690960902719267.
Iatridou, Sabine & Anthony Kroch. 1992. The licensing of CP-recursion and its relevance to the Germanic verb-second phenomenon. *Working Papers in Scandinavian Syntax* 50. 1–24.
Ingham, Richard. 2018. Topic, focus and null subjects in Old French. *Canadian Journal of Linguistics* 63(2). 242–263. doi:10.1017/cnj.2017.48.
Jensen, Isabel Nadine & Marit Westergaard. 2023. Syntax matters: Exploring the effect of linguistic similarity in third language acquisition. *Language Learning* 73(2). 374–402. doi:10.1111/lang.12525.
Jouitteau, Mélanie. 2007. The Brythonic reconciliation. From verb-first to generalized verb-second. *Linguistic Variation Yearbook* 7. 163–200. doi:10.1075/livy.7.06jou.
Julien, Marit. 2015. The force of V2 revisited. *Journal of Comparative Germanic Linguistics* 18(2). 139–181. doi:10.1007/s10828-015-9073-2.
Kaiser, Georg A. 2002–2003. Die Verb-Zweit-Stellung im Rätoromanischen. Ein typologischer Vergleich. *Ladinia* 26–27. 313–334.
Kaiser, Georg A. 2002. *Verbstellung und Verbstellungswandel in den romanischen Sprachen*. Tübingen: Max Niemeyer. doi:10.1515/9783110911640.
Kaiser, Georg A. & Franziska Maria Hack. 2013. Language change in comparison: The (special) case of Raeto-Romance. In Jürg Fleischer & Horst J. Simon (eds.), *Sprachwandelvergleich – Comparing diachronies*, 73–98. Berlin, Boston: De Gruyter. doi:10.1515/9783110313468.73.
Kaiser, Georg A. & Michael Zimmermann. 2011. On the decrease in subject-verb inversion in French declaratives. In Esther Rinke & Tanja Kupisch (eds.), *Development of grammar: Language acquisition and diachronic change. In honour of Jürgen M. Meisel*, 355–381. Amsterdam, Philadelphia: John Benjamins. doi:10.1075/hsm.11.19kai.
Källgren, Gunnel & Ellen F. Prince. 1989. Swedish VP-topicalization and Yiddish verb-topicalization. *Nordic Journal of Linguistics* 12(1). 47–58. doi:10.1017/S033258650000192X.
van Kemenade, Ans. 1987. *Syntactic case and morphological case in the history of English*. Dordrecht: Foris. doi:10.1515/9783110882308.
van Kemenade, Ans. 1997. V2 and embedded topicalization in Old and Middle English. In Ans van Kemenade & Nigel Vincent (eds.), *Parameters of morphosyntactic change*, 326–352. Cambridge, New York, Melbourne: Cambridge University Press.
van Kemenade, Ans. 2012. Rethinking the loss of verb second. In Terttu Nevalainen & Elizabeth Closs Traugott (eds.), *The Oxford handbook of the history of English*, 822–834. New York: Oxford University Press. doi:10.1093/oxfordhb/9780199922765.013.0067.

van Kemenade, Ans & Bettelou Los. 2006. Discourse adverbs and clausal syntax in Old and Middle English. In Ans van Kemenade & Bettelou Los (eds.), *The handbook of the history of English*, 224–248. Malden, MA, Oxford, Carlton: Blackwell. doi:10.1002/9780470757048.ch10.

van Kemenade, Ans & Tanja Mel̀ićev. 2011. Syntax and discourse in Old and Middle English word order. In Dianne Jonas, John Whitman & Andrew Garrett (eds.), *Grammatical change: Origins, nature, outcomes*, 239–255. Oxford, New York: Oxford University Press. doi:10.1093/acprof:oso/9780199582624.003.0012.

van Kemenade, Ans & Marit Westergaard. 2012. Syntax and information structure. Verb-second variation in Middle English. In Anneli Meurman-Solin, Maria Jose Lopez-Couso & Bettelou Los (eds.), *Information structure and syntactic change in the history of English*, 87–118. Oxford, New York: Oxford University Press. doi:10.1093/acprof:oso/9780199860210.003.0005.

Kennard, Holly J. 2018. Non-negative word order in Breton: Maintaining verb-second. *Transactions of the Philological Society* 116(2). 153–178. doi:10.1111/1467-968X.12119.

Keogh, Aislinn, Simon Kirby & Jennifer Culbertson. 2022. In search of a unified mechanism for regularisation of linguistic variation. *Amsterdam Colloquium* 23. 360–366.

Kim, Kathy MinHye & Kimberly M. Fenn. 2020. Sleep-dependent consolidation of second language grammar knowledge: The role of awareness. *Studies in Second Language Acquisition* 42(5). 1107–1120. doi:10.1017/S0272263120000200.

Kim, Kathy Minhye & Aline Godfroid. 2019. Should we listen or read? Modality effects in implicit and explicit knowledge. *The Modern Language Journal* 103(3). 648–664. doi:10.1111/modl.12583.

Kiparsky, Paul. 1995. Indo-European origins of Germanic syntax. In Adrian Battye & Ian Roberts (eds.), *Clause structure and language change*, 140–169. New York, Oxford: Oxford University Press. doi:10.1093/oso/9780195086324.003.0006.

Kirby, Simon, Hannah Cornish & Kenny Smith. 2008. Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences* 105(31). 10681–10686. doi:10.1073/pnas.0707835105.

Kirby, Simon, Monica Tamariz, Hannah Cornish & Kenny Smith. 2015. Compression and communication in the cultural evolution of linguistic structure. *Cognition* 141. 87–102. doi:10.1016/j.cognition.2015.03.016.

Kirton, Fiona, Simon Kirby, Kenny Smith, Jennifer Culbertson & Marieke Schouwstra. 2021. Constituent order in silent gesture reflects the perspective of the producer. *Journal of Language Evolution* 6(1). 54–76. doi:10.1093/jole/lzaa010.

Klævik-Pettersen, Espen. 2019. What is Germanic and what is not about Old French V2. *Linguistic Variation* 19(1). 141–198. doi:10.1075/lv.00011.kla.

de Kok, Daniël & Sebastian Pütz. 2019. *TüBa-D/DP stylebook*. University of Tübingen: Seminar für Sprachwissenschaft.

Koopman, Hilda & Dominique Sportiche. 1991. The position of subjects. *Lingua* 85(2). 211–258. doi:10.1016/0024-3841(91)90022-W.

von Koss Torkildsen, Janne, Natalie S. Dailey, Jessica M. Aguilar, Rebecca Gómez & Elena Plante. 2013. Exemplar variability facilitates rapid learning of an otherwise unlearnable grammar by individuals with language-based learning disability. *Journal of Speech, Language, and Hearing Research* 56(2). 618–629. doi:10.1044/1092-4388(2012/11-0125).
Kroch, Anthony. 1989. Reflexes of grammar in patterns of language change. *Language Variation and Change* 1(3). 199–244. doi:10.1017/S0954394500000168.
Kroch, Anthony. 1994. Morphosyntactic variation. *Chicago Linguistic Society (CLS)* 30(2). 180–201.
Kroch, Anthony & Ann Taylor. 1997. Verb movement in Old and Middle English: Dialect variation and language contact. In Ans van Kemenade & Nigel Vincent (eds.), *Parameters of morphosyntactic change*, 297–325. Cambridge, New York, Melbourne: Cambridge University Press.
Kroch, Anthony, Ann Taylor & Donald Ringe. 2000. The Middle English verb-second constraint: A case study in language contact and language change. In Susan C. Herring, Pieter van Reenen & Lene Schøsler (eds.), *Textual parameters in older languages*, 353–391. Amsterdam, Philadelphia: John Benjamins. doi:10.1075/cilt.195.17kro.
Kurumada, Chigusa, Stephan C. Meylan & Michael C. Frank. 2013. Zipfian frequency distributions facilitate word segmentation in context. *Cognition* 127(3). 439–453. doi:10.1016/j.cognition.2013.02.002.
Labelle, Marie & Paul Hirschbühler. 2018. Topic and focus in Old French V1 and V2 structures. *Canadian Journal of Linguistics* 63(2). 264–287. doi:10.1017/cnj.2017.52.
Larrivée, Pierre. 2019. To be or not to be informational: Preverbal complements in Medieval French V2 configurations. *Glossa: a journal of general linguistics* 4(1). 1–19. doi:10.5334/gjgl.851.
Larrivée, Pierre. 2021. An information structure scenario for V2 loss in Medieval French. *Diachronica* 38(2). 189–209. doi:10.1075/dia.19040.lar.
Larrivée, Pierre. 2022. Is Medieval French diglossic? New evidence on remnant V2 and register. *Isogloss. Open Journal of Romance Linguistics* 8(2). 1–16. doi:10.5565/rev/isogloss.114.
Larsson, Ida & Kari Kinn. 2022. Stability and change in the C-domain in American Swedish. *Languages* 7(4). 1–22. doi:10.3390/languages7040256.
Lavi-Rotbain, Ori & Inbal Arnon. 2021. Visual statistical learning is facilitated in Zipfian distributions. *Cognition* 206. 1–9. doi:10.1016/j.cognition.2020.104492.
Lavi-Rotbain, Ori & Inbal Arnon. 2022. The learnability consequences of Zipfian distributions in language. *Cognition* 223. 1–15. doi:10.1016/j.cognition.2022.105038.
Lavi-Rotbain, Ori & Inbal Arnon. 2023. Zipfian distributions in child-directed speech. *Open Mind* 7. 1–30. doi:10.1162/opmi_a_00070.
Ledgeway, Adam. 2017. Late Latin verb second: The sentential word order of the Itinerarium Egeriae. *Catalan Journal of Linguistics* 16. 163–216. doi:10.5565/rev/catjl.212.
Lee, Juyoung & Jeffrey Witzel. 2023. Plausibility and structural reanalysis in L1 and L2 sentence comprehension. *Quarterly Journal of Experimental Psychology* 76(2). 319–337. doi:10.1177/17470218221092400.

de Leeuw, Joshua R., Rebecca A. Gilbert & Björn Luchterhandt. 2023. jsPsych: Enabling an open-source collaborative ecosystem of behavioral experiments. *Journal of Open Source Software* 8(85). 1–4. doi:10.21105/joss.05351.

Leong, Christine Xiang Ru, Jessica M. Price, Nicola J. Pitchford & Walter J. B. van Heuven. 2018. High variability phonetic training in adaptive adverse conditions is rapid, effective, and sustained. *PLoS ONE* 13(10). 1–24. doi:10.1371/journal.pone.0204888.

Light, Caitlin. 2012. *The syntax and pragmatics of fronting in Germanic*. Philadelphia, PA: University of Pennsylvania dissertation.

Lightfoot, David. 1994. Degree-0 learnability. In Barbara Lust, Gabriella Hermon & Jaklin Kornfilt (eds.), *Syntactic theory and first language acquisition: Cross-linguistic perspectives*, 453–471. Hillsdale, NJ, Hove: Lawrence Erlbaum Associates. doi:10.4324/9781315789200.

Lightfoot, David. 1995. Why UG needs a learning theory: Triggering verb movement. In Adrian Battye & Ian Roberts (eds.), *Clause structure and language change Structure and Language Change*, 31–52. New York, Oxford: Oxford University Press. doi:10.1093/oso/9780195086324.003.0002.

Lightfoot, David. 1999. *The development of language. Acquisition, change, and evolution*. Malden, MA: Blackwell.

Lightfoot, David. 2006. *How new languages emerge*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511616204.

Lightfoot, David & Marit Westergaard. 2007. Language acquisition and language change: Inter-relationships. *Language and Linguistics Compass* 1(5). 396–415. doi:10.1111/j.1749-818X.2007.00023.x.

Liter, Adam, Christopher C. Heffner & Cristina Schmitt. 2017. The interpretation of plural morphology and (non-)obligatory number marking: An argument from artificial language learning. *Language Learning and Development* 13(4). 451–480. doi:10.1080/15475441.2017.1324307.

Lively, Scott E., John S. Logan & David B. Pisoni. 1993. Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *The Journal of the Acoustical Society of America* 94(3). 1242–1255. doi:10.1121/1.408177.

Liver, Ricarda. 2010. *Rätoromanisch: Eine Einführung in das Bündnerromanische*. 2nd edn. Tübingen: Narr Francke Attempto.

Lohndal, Terje, Marit Westergaard & Øystein A. Vangsnes. 2020. Verb second in Norwegian: Variation and acquisition. In Rebecca Woods & Sam Wolfe (eds.), *Rethinking verb second*, 770–789. Oxford: Oxford University Press. doi:10.1093/oso/9780198844303.003.0033.

Los, Bettelou. 2009. The consequences of the loss of verb-second in English: Information structure and syntax in interaction. *English Language and Linguistics* 13(1). 97–125. doi:10.1017/S1360674308002876.

Los, Bettelou. 2012. The loss of verb-second and the switch from bounded to unbounded systems. In Anneli Meurman-Solin, Maria Jose Lopez-Couso & Bettelou Los (eds.),

Information structure and syntactic change in the history of English, 21–46. Oxford, New York: Oxford University Press. doi:10.1093/acprof:oso/9780199860210.003.0002.
Los, Bettelou & Gea Dreschler. 2012. The loss of local anchoring: From adverbial local anchors to permissive subjects. In Terttu Nevalainen & Elizabeth Closs Traugott (eds.), *The Oxford handbook of the history of English*, 859–872. New York: Oxford University Press. doi:10.1093/oxfordhb/9780199922765.013.0070.
Los, Bettelou & Ans van Kemenade. 2018. Syntax and the morphology of deixis. The loss of demonstratives and paratactic clause linking. In Marco Coniglio, Andrew Murphy, Eva Schlachter & Tonjes Veenstra (eds.), *Atypical demonstratives, syntax, semantics and pragmatics*, 127–158. Berlin, Boston: De Gruyter. doi:10.1515/9783110560299-005.
Lundquist, Björn, Merete Anderssen, Terje Lohndal & Marit Westergaard. 2020. Variation across individuals and domains in Norwegian heritage language. *Oslo Studies in Language* 11(2). 251–267. doi:10.5617/osla.8502.
MacWhinney, Brian. 2000. *The CHILDES project: Tools for analyzing talk*. 3rd edn. Mahwah, NJ: Lawrence Erlbaum Associates. doi:10.4324/9781315805672.
Maldonado, Mora & Jennifer Culbertson. 2022. Person of interest: Experimental investigations into the learnability of person systems. *Linguistic Inquiry* 53(2). 295–336. doi:10.1162/ling_a_00406.
Malik-Moraleda, Saima, Maya Taliaferro, Steve Shannon, Niharika Jhingan, Sara Swords, David J. Peterson, Paul Frommer, Marc Okrand, Jessie Sams, Ramsey Cardwell, Cassie Freeman & Evelina Fedorenko. 2023. Constructed languages are processed by the same brain mechanisms as natural languages. bioRxiv. doi:10.1101/2023.07.28.550667.
Manetta, Emily. 2011. *Peripheries in Kashmiri and Hindi-Urdu. The syntax of discourse-driven movement*. Amsterdam, Philadelphia: John Benjamins. doi:10.1075/lfab.4.
Manetta, Emily. 2021. Verb-second and the verb-stranding verb phrase ellipsis debate. *Glossa: a journal of general linguistics* 6(1). 1–35. doi:10.16995/glossa.5704.
Martin, Alexander, David Adger, Klaus Abels, Patrick Kanampiu & Jennifer Culbertson. 2024. A universal cognitive bias in word order: Evidence from speakers whose language goes against it. *Psychological Science* 35(3). 304–311. doi:10.1177/09567976231222836.
Martin, Alexander & Jennifer Culbertson. 2020. Revisiting the suffixing preference: Native-language affixation patterns influence perception of sequences. *Psychological Science* 31(9). 1107–1116. doi:10.1177/0956797620931108.
Martin, Alexander, Annie Holtz, Klaus Abels, David Adger & Jennifer Culbertson. 2020. Experimental evidence for the influence of structure and meaning on linear order in the noun phrase. *Glossa: a journal of general linguistics* 5(1). 1–21. doi:10.5334/gjgl.1085.
Martin, Alexander, Theeraporn Ratitamkul, Klaus Abels, David Adger & Jennifer Culbertson. 2019. Cross-linguistic evidence for cognitive universals in the noun phrase. *Linguistics Vanguard* 5(1). 1–13. doi:10.1515/lingvan-2018-0072.
Martins, Ana Maria. 2019. Against V2 as a general property of Old Romance languages. In Ingo Feldhausen, Martin Elsig, Imme Kuchenbrandt & Marieke Neuhaus (eds.), *Romance Languages and Linguistic Theory 15. Selected papers from 'Going Romance' 30,*

Frankfurt, 9–33. Amsterdam, Philadelphia: John Benjamins. doi:10.1075/rllt.15.01mar.
Maurer-Cecchini, Philippe. 2021. *A grammar of Tuatschin. A Sursilvan Romansh dialect*. Berlin: Language Science Press. doi:10.5281/zenodo.5137647.
Meelen, Marieke. 2016. *Why Jesus and Job spoke bad Welsh. The origin and distribution of V2 orders in Middle Welsh*. Leiden: Universiteit Leiden Proefschrift.
Meelen, Marieke. 2020. Reconstructing the rise of verb second in Welsh. In Rebecca Woods & Sam Wolfe (eds.), *Rethinking verb second*, 426–454. Oxford: Oxford University Press. doi:10.1093/oso/9780198844303.003.0018.
Meelen, Marieke, Khalid Mourigh & Lisa Lai-Shen Cheng. 2020. V3 in urban youth varieties of Dutch. In András Bárány, Theresa Biberauer, Jamie Douglas & Sten Vikner (eds.), *Syntactic architecture and its consequences II: Between syntax and morphology*, 327–355. Berlin: Language Science Press. doi:10.5281/zenodo.4280657.
Meinunger, André. 2007. About object es in the German vorfeld. *Linguistic Inquiry* 38(3). 553–563. doi:10.1162/ling.2007.38.3.553.
Meisel, Jürgen M. 1997. The acquisition of the syntax of negation in French and German: Contrasting first and second language development. *Second Language Research* 13(3). 227–263. doi:10.1191/026765897666180760.
Meisezahl, Marc. 2019. *Word order in subordinated clauses in the Surselva*. Konstanz: Universität Konstanz Masterarbeit.
Miller, Zachary F. & Aline Godfroid. 2020. Emotions in incidental language learning. An individual differences approach. *Studies in Second Language Acquisition* 42(1). 115–141. doi:10.1017/S027226311900041X.
Mitrofanova, Natalia, Evelina Leivada & Marit Westergaard. 2023. Crosslinguistic influence in L3 acquisition: Evidence from artificial language learning. *Linguistic Approaches to Bilingualism* 13(5). 717–742. doi:10.1075/lab.22063.mit.
Miyashita, Mizuki. 2006. Tohono O'odham. In Keith Brown (ed.), *Encyclopedia of Language & Linguistics*, 2nd, 735–737. Amsterdam, Boston: Elsevier. doi:10.1016/B0-08-044854-2/05047-1.
Mohr, Sabine. 2009. V2 as a single-edge phenomenon. In Kleanthes K. Grohmann & Phoevos Panagiotidis (eds.), *Selected papers from the 2006 Cyprus syntaxfest*, 141–158. Newcastle upon Tyne: Cambridge Scholars.
Moreton, Elliott. 2008. Analytic bias and phonological typology. *Phonology* 25(1). 83–127. doi:10.1017/S0952675708001413.
Morgan, James L., Richard P. Meier & Elissa L. Newport. 1987. Structural packaging in the input to language learning: Contributions of prosodic and morphological marking of phrases to the acquisition of language. *Cognitive Psychology* 19(4). 498–550. doi:10.1016/0010-0285(87)90017-X.
Morgan, James L., Richard P. Meier & Elissa L. Newport. 1989. Facilitating the acquisition of syntax with cross-sentential cues to phrase structure. *Journal of Memory and Language* 28(3). 360–374. doi:10.1016/0749-596X(89)90039-9.
Morgan-Short, Kara, Karsten Steinhauer, Cristina Sanz & Michael T. Ullman. 2012. Explicit and implicit second language training differentially affect the achievement of

native-like brain activation patterns. *Journal of Cognitive Neuroscience* 24(4). 933–947. doi:10.1162/jocn_a_00119.
Moss, Aaron J., Cheskie Rosenzweig, Jonathan Robinson & Leib Litman. 2020. Demographic stability on Mechanical Turk despite COVID-19. *Trends in Cognitive Sciences* 24(9). 678–680. doi:10.1016/j.tics.2020.05.014.
Motamedi, Yasamin, Marieke Schouwstra, Kenny Smith, Jennifer Culbertson & Simon Kirby. 2019. Evolving artificial sign languages in the lab: From improvised gesture to systematic sign. *Cognition* 192. 1–21. doi:10.1016/j.cognition.2019.05.001.
Müller, Gereon. 2004. Verb-second as vP-first. *Journal of Comparative Germanic Linguistics* 7(3). 179–234. doi:10.1023/B:JCOM.0000016453.71478.3a.
Naigles, Letitia. 1990. Children use syntax to learn verb meanings. *Journal of Child Language* 17(2). 357–374. doi:10.1017/S0305000900013817.
Napoli, Donna Jo & Rachel Sutton–Spence. 2021. Clause-initial Vs in sign languages: Scene-setters. In Vera Lee-Schoenfeld & Dennis Ott (eds.), *Parameters of predicate fronting*, 192–219. Oxford: Oxford University Press. doi:10.1093/oso/9780197545553.003.0008.
Önnerfors, Olaf. 1997. On narrative declarative V1 sentences in German. In Toril Swan & Olaf Jansen Westvik (eds.), *Modality in Germanic languages. Historical and comparative perspectives*, 293–320. Berlin, New York: Mouton de Gruyter. doi:10.1515/9783110889932.293.
Onnis, Luca & Erik Thiessen. 2013. Language experience changes subsequent learning. *Cognition* 126(2). 268–284. doi:10.1016/j.cognition.2012.10.008.
Pagán, Ascensión & Kate Nation. 2019. Learning words via reading: Contextual diversity, spacing, and retrieval effects in adults. *Cognitive Science* 43(1). 1–24. doi:10.1111/cogs.12705.
Peer, Eyal, David Rothschild, Andrew Gordon, Zak Evernden & Ekaterina Damer. 2021. Data quality of platforms and panels for online behavioral research. *Behavior Research Methods* 54(4). 1643–1662. doi:10.3758/s13428-021-01694-3.
Perkins, Laurel & Jeffrey Lidz. 2021. Eighteen-month-old infants represent nonlocal syntactic dependencies. *Proceedings of the National Academy of Sciences* 118(41). 1–7. doi:10.1073/pnas.2026469118.
Pintzuk, Susan. 1991. *Phrase structures in competition: Variation and change in Old English word order*. Philadelphia, PA: University of Pennsylvania dissertation.
Pintzuk, Susan & Eric Haeberli. 2008. Structural variation in Old English root clauses. *Language Variation and Change* 20(3). 367–407. doi:10.1017/S095439450800015X.
Poletiek, Fenna H. & Tessa J. P. van Schijndel. 2009. Stimulus set size and statistical coverage of the grammar in artificial grammar learning. *Psychonomic Bulletin & Review* 16(6). 1058–1064. doi:10.3758/PBR.16.6.1058.
Poletto, Cecilia. 2000. *The higher functional field: Evidence from Northern Italian dialects*. New York, Oxford: Oxford University Press. doi:10.1093/oso/9780195133561.001.0001.
Poletto, Cecilia. 2002. The left-periphery of V2-Rhaetoromance dialects: A new view on V2 and V3. In Sjef Barbiers, Leonie Cornips & Susanne van der Kleij (eds.), *Syntactic microvariation*, 214–242. Amsterdam: Meertens Institute.

Poletto, Cecilia. 2014. *Word order in Old Italian*. Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780199660247.001.0001.

Poletto, Cecilia. 2016. Which clues for which V2. A contribution to the typology of V2 on the basis of Old Italian. In Ermenegildo Bidese, Federica Cognola & Manuela Caterina Moroni (eds.), *Theoretical approaches to linguistic variation*, 237–258. Amsterdam, Philadelphia: John Benjamins. doi:10.1075/la.234.09pol.

Polišenská, Kamila, Shula Chiat, Jakub Szewczyk & Katherine E. Twomey. 2021. Effects of semantic plausibility, syntactic complexity and n-gram frequency on children's sentence repetition. *Journal of Child Language* 48(2). 261–284. doi:10.1017/S0305000920000306.

Postal, Paul Martin. 2004. *Skeptical linguistic essays*. Oxford, New York: Oxford University Press. doi:10.1093/oso/9780195166712.001.0001.

Prickett, Brandon. 2019. Learning biases in opaque interactions. *Phonology* 36(4). 627–653. doi:10.1017/S0952675719000320.

Puggaard, Rasmus. 2019. Flexibility of frequent clause openers in talk-in-interaction: Det 'it, that' and så 'then' in the prefield in Danish. *Nordic Journal of Linguistics* 42(3). 291–327. doi:10.1017/S0332586519000088.

Pujol i Campeny, Afra. 2018. *Word order in Old Catalan*. Cambridge: University of Cambridge dissertation. doi:10.17863/CAM.40198.

R Core Team. 2020. *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

R Core Team. 2022. *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

Rankin, Tom. 2014. Word order and case in the comprehension of L2 German by L1 English speakers. *EUROSLA Yearbook* 14. 201–224. doi:10.1075/eurosla.14.08ran.

Raviv, Limor, Gary Lupyan & Shawn C. Green. 2022. How variability shapes learning and generalization. *Trends in Cognitive Sciences* 26(6). 462–483. doi:10.1016/j.tics.2022.03.007.

Reber, Arthur S. 1967. Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior* 6(6). 855–863. doi:10.1016/S0022-5371(67)80149-X.

Rebuschat, Patrick. 2008. *Implicit learning of natural language syntax*. Cambridge: University of Cambridge dissertation. doi:10.17863/CAM.15883.

Rebuschat, Patrick & John N. Williams. 2012. Implicit and explicit knowledge in second language acquisition. *Applied Psycholinguistics* 33(4). 829–856. doi:10.1017/S0142716411000580.

Reeder, Patricia A., Elissa L. Newport & Richard N. Aslin. 2013. From shared contexts to syntactic categories: The role of distributional information in learning linguistic form-classes. *Cognitive Psychology* 66(1). 30–54. doi:10.1016/j.cogpsych.2012.09.001.

Reeder, Patricia A., Elissa L. Newport & Richard N. Aslin. 2017. Distributional learning of subcategories in an artificial grammar: Category generalization and subcategory restrictions. *Journal of Memory and Language* 97. 17–29. doi:10.1016/j.jml.2017.07.006.

Ribeiro, Ilza. 1995. Evidence for a verb-second phase in Old Portuguese. In Adrian Battye & Ian Roberts (eds.), *Clause structure and language change*, 110–139. New York, Oxford: Oxford University Press. doi:10.1093/oso/9780195086324.003.0005.
Rinke, Esther. 2009. Verb placement in Old Portuguese. In Daniel Jakob & Andreas Dufner (eds.), *Focus and background in Romance languages*, 309–332. Amsterdam, Philadelphia: John Benjamins. doi:10.1075/slcs.112.11rin.
Rizzi, Luigi. 1996. Residual verb second and the wh-criterion. In Adriana Belletti & Luigi Rizzi (eds.), *Parameters and functional heads. Essays in comparative syntax*, 63–90. New York, Oxford: Oxford University Press. doi:10.1093/oso/9780195087932.003.0003.
Rizzi, Luigi. 1997. The fine structure of the left periphery. In Liliane Haegeman (ed.), *Elements of grammar. Handbook in generative syntax*, 281–337. Dordrecht, Boston, London: Kluwer. doi:10.1007/978-94-011-5420-8_7.
Roberts, Gareth & Maryia Fedzechkina. 2018. Social biases modulate the loss of redundant forms in the cultural evolution of language. *Cognition* 171. 194–201. doi:10.1016/j.cognition.2017.11.005.
Roberts, Ian. 1993. *Verbs and diachronic syntax. A comparative history of English and French*. Dordrecht, Boston, London: Kluwer. doi:10.1007/978-94-011-2910-7.
Roberts, Ian. 1996. Remarks on the Old English C-system and the diachrony of V2. In Ellen Brandner & Gisella Ferraresi (eds.), *Language change and generative grammar*, 154–167. Opladen: Westdeutscher Verlag. doi:10.1007/978-3-322-90776-9_6.
Roberts, Ian. 2004. The C-System in Brythonic Celtic languages, V2, and the EPP. In Luigi Rizzi (ed.), *The structure of CP and IP: The cartography of syntactic structures*, 297–328. Oxford, New York: Oxford University Press. doi:10.1093/oso/9780195159486.003.0010.
Roberts, Ian. 2012. Phases, head movement and second-position effects. In Ángel J. Gallego (ed.), *Phases. Developing the framework*, 385–440. Berlin, Boston: De Gruyter Mouton. doi:10.1515/9783110264104.385.
Roberts, Leah & Claudia Felser. 2011. Plausibility and recovery from garden paths in second language sentence processing. *Applied Psycholinguistics* 32(2). 299–331. doi:10.1017/S0142716410000421.
Roeper, Thomas. 1999. Universal bilingualism. *Bilingualism: Language and Cognition* 2(3). 169–186. doi:10.1017/S1366728999000310.
Rost, Gwyneth C. & Bob McMurray. 2010. Finding the signal by adding noise: The role of noncontrastive phonetic variability in early word learning: *Infancy* 15(6). 608–635. doi:10.1111/j.1532-7078.2010.00033.x.
Rudanko, Juhani. 1982. Towards a description of negatively conditioned subject operator inversion in English. *English Studies* 63(4). 348–359. doi:10.1080/00138388208598192.
Ruiz, Simón, Kaitlyn M. Tagarelli & Patrick Rebuschat. 2018. Simultaneous acquisition of words and syntax: Effects of exposure condition and declarative memory. *Frontiers in Psychology* 9. 1–11. doi:10.3389/fpsyg.2018.01168.
Saffran, Jenny R., Richard N. Aslin & Elissa L. Newport. 1996. Statistical learning by 8-month-old infants. *Science* 274(5294). 1926–1928. doi:10.1126/science.274.5294.1926.

Sahkai, Heete & Anne Tamm. 2019. Verb placement and accentuation: Does prosody constrain the Estonian V2? *Open Linguistics* 5(1). 729–753. doi:10.1515/opli-2019-0040.

Sailor, Craig. 2020. Rethinking 'residual' verb second. In Rebecca Woods & Sam Wolfe (eds.), *Rethinking verb second*, 126–149. Oxford: Oxford University Press. doi:10.1093/oso/9780198844303.003.0006.

Saldana, Carmen, Borja Herce & Balthasar Bickel. 2022. More or less unnatural: Semantic similarity shapes the learnability and cross-linguistic distribution of unnatural syncretism in morphological paradigms. *Open Mind* 6. 183–210. doi:10.1162/opmi_a_00062.

Saldana, Carmen, Simon Kirby, Robert Truswell & Kenny Smith. 2019. Compositional hierarchical structure evolves through cultural transmission: An experimental study. *Journal of Language Evolution* 4(2). 83–107. doi:10.1093/jole/lzz002.

Saldana, Carmen, Yohei Oseki & Jennifer Culbertson. 2021. Cross-linguistic patterns of morpheme order reflect cognitive biases: An experimental study of case and number morphology. *Journal of Memory and Language* 118. 1–13. doi:10.1016/j.jml.2020.104204.

Saldana, Carmen, Kenny Smith, Simon Kirby & Jennifer Culbertson. 2021. Is regularization uniform across linguistic levels? Comparing learning and production of unconditioned probabilistic variation in morphology and word order. *Language Learning and Development* 17(2). 158–188. doi:10.1080/15475441.2021.1876697.

Salvesen, Christine Meklenborg & George Walkden. 2017. Diagnosing embedded V2 in Old English and Old French. In Eric Mathieu & Robert Truswell (eds.), *Micro-change and macro-change in diachronic syntax*, 168–181. Oxford: Oxford University Press. doi:10.1093/oso/9780198747840.003.0011.

Santorini, Beatrice. 1992. Variation and change in Yiddish subordinate clause word order. *Natural Language and Linguistic Theory* 10(4). 595–640. doi:10.1007/BF00133331.

Saratsli, Dionysia, Stefan Bartell & Anna Papafragou. 2020. Cross-linguistic frequency and the learnability of semantics: Artificial language learning studies of evidentiality. *Cognition* 197. 1–16. doi:10.1016/j.cognition.2020.104194.

Saratsli, Dionysia & Anna Papafragou. 2023. Pragmatic effects on semantic learnability: Insights from evidentiality. *Journal of Memory and Language* 131. 1–16. doi:10.1016/j.jml.2023.104426.

Schafer, Robin. 1995. Negation and verb second in Breton. *Natural Language and Linguistic Theory* 13(1). 135–172. doi:10.1007/BF00992780.

Schäfer, Lisa. 2021. Topic drop in German: Empirical support for an information-theoretic account to a long-known omission phenomenon. *Zeitschrift für Sprachwissenschaft* 40(2). 161–197. doi:10.1515/zfs-2021-2024.

Schiff, Rachel, Pesi Ashkenazi, Shani Kahta & Ayelet Sasson. 2021. Stimulus variation-based training enhances artificial grammar learning. *Acta Psychologica* 214. 1–9. doi:10.1016/j.actpsy.2021.103252.

Schnoebelen, Tyler & Victor Kuperman. 2010. Using Amazon Mechanical Turk for linguistic research. *Psihologija* 43(4). 441–464. doi:10.2298/PSI1004441S.

Schouwstra, Marieke & Henriëtte de Swart. 2014. The semantic origins of word order. *Cognition* 131(3). 431–436. doi:10.1016/j.cognition.2014.03.004.
Schütze, Carson T. 2016. *The empirical base of linguistics. Grammaticality judgements and linguistic methodology*. Berlin: Language Science Press. doi:10.5281/ZENODO.3362734.
Schwartz, Bonnie D. & Rex A. Sprouse. 1996. L2 cognitive states and the Full Transfer/Full Access model. *Second Language Research* 12(1). 40–72. doi:10.1177/026765839601200103.
Schwartz, Bonnie D. & Sten Vikner. 1989. All verb second clauses are CPs. *Working Papers in Scandinavian Syntax* 43. 27–49.
Schwartz, Bonnie D. & Sten Vikner. 1996. The verb always leaves IP in V2 clauses. In Adriana Belletti & Luigi Rizzi (eds.), *Parameters and functional heads. Essays in comparative syntax*, 11–62. New York, Oxford: Oxford University Press. doi:10.1093/oso/9780195087932.003.0002.
Shannon, Claude E. 1948. A mathematical theory of communication. *Bell System Technical Journal* 27(3). 379–423. doi:10.1002/j.1538-7305.1948.tb01338.x.
Shapiro, Naomi Tachikawa & Shane Steinert-Threlkeld. 2023. Iconic artificial language learning: A conceptual replication with English speakers. *Cognitive Science Society (CSS)* 45. 146–152.
Siegelman, Noam, Louisa Bogaerts, Amit Elazar, Joanne Arciuli & Ram Frost. 2018. Linguistic entrenchment: Prior knowledge impacts statistical learning performance. *Cognition* 177. 198–213. doi:10.1016/j.cognition.2018.04.011.
Silvey, Catriona, Simon Kirby & Kenny Smith. 2015. Word meanings evolve to selectively preserve distinctions on salient dimensions. *Cognitive Science* 39(1). 212–226. doi:10.1111/cogs.12150.
Singh, Leher. 2008. Influences of high and low variability on infant word recognition. *Cognition* 106(2). 833–870. doi:10.1016/j.cognition.2007.05.002.
Singh, Onkar. 2021. Verb second in Old Venetian. *Isogloss. Open Journal of Romance Linguistics* 7(7). 1–28. doi:10.5565/rev/isogloss.110.
Sinkeviciute, Ruta, Helen Brown, Gwen Brekelmans & Elizabeth Wonnacott. 2019. The role of input variability and learner age in second language vocabulary learning. *Studies in Second Language Acquisition* 41(4). 795–820. doi:10.1017/S0272263119000263.
Sitaridou, Ioanna. 2011. Word order and information structure in Old Spanish. *Catalan Journal of Linguistics* 10(1). 159–184. doi:10.5565/rev/catjl.36.
Sitaridou, Ioanna. 2012. A comparative study of word order in Old Romance. *Folia Linguistica* 46(2). 553–604. doi:10.1515/flin.2012.019.
Smith, Kenny & Elizabeth Wonnacott. 2010. Eliminating unpredictable variation through iterated learning. *Cognition* 116(3). 444–449. doi:10.1016/j.cognition.2010.06.004.
Sneller, Betsy & Gareth Roberts. 2018. Why some behaviors spread while others don't: A laboratory simulation of dialect contact. *Cognition* 170. 298–311. doi:10.1016/j.cognition.2017.10.014.
Sommers, Mitchell S. & Joe Barcroft. 2006. Stimulus variability and the phonetic relevance hypothesis: Effects of variability in speaking style, fundamental frequency,

and speaking rate on spoken word identification. *The Journal of the Acoustical Society of America* 119(4). 2406–2416. doi:10.1121/1.2171836.
Spescha, Arnold. 1989. *Grammatica sursilvana*. Cuera: Casa editura per mieds d'instrucziun.
Speyer, Augustin. 2010. *Topicalization and stress clash avoidance in the history of English*. Berlin, New York: De Gruyter Mouton. doi:10.1515/9783110220247.
Speyer, Augustin. 2012. Stress clash and word order changes in the left periphery in Old English and Middle English. In Terttu Nevalainen & Elizabeth Closs Traugott (eds.), *The Oxford handbook of the history of English*, 873–883. New York: Oxford University Press. doi:10.1093/oxfordhb/9780199922765.013.0071.
Sprouse, Jon. 2011. A validation of Amazon Mechanical Turk for the collection of acceptability judgments in linguistic theory. *Behavior Research Methods* 43(1). 155–167. doi:10.3758/s13428-010-0039-7.
Steiner, Brittany Devan Jelm. 2014. *The evolution of information structure and verb second in the history of French*. Bloomington: Indiana University dissertation.
Storto, Luciana R. 1999. *Aspects of a Karitiana grammar*. Cambridge, MA: MIT dissertation.
Storto, Luciana R. 2003. Interactions between verb movement and agreement in Karitiana (Tupi Stock). *Revista Letras* 60. 411–433. doi:10.5380/rel.v60i0.2876.
Storto, Luciana R. 2020. Arguments for and against verb second in Karitiana. *Revista Linguíſtica* 16. 316–335. doi:10.31513/linguistica.2020.v16nEsp.a21873.
Sumner, Meghan. 2011. The role of variation in the perception of accented speech. *Cognition* 119(1). 131–136. doi:10.1016/j.cognition.2010.10.018.
de Swart, Peter & Geertje van Bergen. 2019. How animacy and verbal information influence V2 sentence processing: Evidence from eye movements. *Open Linguistics* 5(1). 630–649. doi:10.1515/opli-2019-0035.
Tabullo, Ángel, Mariana Arismendi, Alejandro Wainselboim, Gerardo Primero, Sergio Vernis, Enrique Segura, Silvano Zanutto & Alberto Yorio. 2012. On the learnability of frequent and infrequent word orders: An artificial language learning study. *Quarterly Journal of Experimental Psychology* 65(9). 1848–1863. doi:10.1080/17470218.2012.677848.
Tagarelli, Kaitlyn M., Mailce Borges Mota & Patrick Rebuschat. 2015. Working memory, learning conditions and the acquisition of L2 syntax. In Zhisheng (Edward) Wen, Mailce Borges Mota & Arthur McNeill (eds.), *Working memory in second language acquisition and processing*, 224–247. Bristol, Buffalo, Toronto: Multilingual Matters. doi:10.21832/9781783093595-017.
Tagarelli, Kaitlyn M., Simón Ruiz, José Luis Moreno Vega & Patrick Rebuschat. 2016. Variability in second language learning. *Studies in Second Language Acquisition* 38(2). 293–316. doi:10.1017/S0272263116000036.
Tal, Shira & Inbal Arnon. 2022. Redundancy can benefit learning: Evidence from word order and case marking. *Cognition* 224. 1–22. doi:10.1016/j.cognition.2022.105055.

Tal, Shira, Kenny Smith, Jennifer Culbertson, Eitan Grossman & Inbal Arnon. 2022. The impact of information structure on the emergence of differential object marking: An experimental study. *Cognitive Science* 46(3). 1–31. doi:10.1111/cogs.13119.
Tang, Kevin & Dinah Baer-Henney. 2023. Modelling L1 and the artificial language during artificial language learning. *Laboratory Phonology* 14(1). 1–55. doi:10.16995/labphon.6460.
Thompson, Susan P. & Elissa L. Newport. 2007. Statistical learning of syntax: The role of transitional probability. *Language Learning and Development* 3(1). 1–42. doi:10.1207/s15473341lld0301_1.
Thráinsson, Höskuldur. 2007. *The syntax of Icelandic*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511619441.
Tily, Harry J., Michael C. Frank & T. Florian Jaeger. 2011. The learnability of constructed languages reflects typological patterns. *Cognitive Science Society (CSS)* 33. 1364–1369.
Travis, Lisa deMena. 1984. *Parameters and effects of word order variation*. Cambridge, MA: MIT dissertation.
Travis, Lisa deMena. 1991. Parameters of phrase structure and verb-second phenomena. In Robert Freidin (ed.), *Principles and parameters in comparative grammar*, 339–364. Cambridge, MA, London: MIT Press.
Traxler, Matthew J. & Martin J. Pickering. 1996. Plausibility and the processing of unbounded dependencies: An eye-tracking study. *Journal of Memory and Language* 35(3). 454–475. doi:10.1006/jmla.1996.0025.
Truswell, Robert. 2021. Grammar competition and word order in a northern early Middle English text. *Languages* 6(2). 1–29. doi:10.3390/languages6020059.
Trutkowski, Ewa. 2016. *Topic drop and null subjects in German*. Berlin, Boston: De Gruyter. doi:10.1515/9783110446173.
Uittenhove, Kim, Stephanie Jeanneret & Evie Vergauwe. 2023. From lab-testing to web-testing in cognitive research: Who you test is more important than how you test. *Journal of Cognition* 6(1). 1–17. doi:10.5334/joc.259.
van Urk, Coppe. 2015. *A uniform syntax for phrasal movement: A case study of Dinka Bor*. Cambridge, MA: Massachusetts Institute of Technology dissertation.
Valian, Virginia, Sandeep Prasada & Jodi Scarpa. 2006. Direct object predictability: Effects on young children's imitation of sentences. *Journal of Child Language* 33(2). 247–269. doi:10.1017/S0305000906007392.
Vance, Barbara. 1995. On the decline of verb movement to comp in Old and Middle French. In Adrian Battye & Ian Roberts (eds.), *Clause structure and language change*, 173–199. New York, Oxford: Oxford University Press. doi:10.1093/oso/9780195086324.003.0007.
Vance, Barbara. 1997. *Syntactic change in Medieval French. Verb-second and null subjects*. Dordrecht: Kluwer. doi:10.1007/978-94-015-8843-0.
Varga, Eva. 2017. *Verbstellungsmuster im Altfranzösischen. Ein Beitrag zur historischen Syntaxforschung aus diskurstraditioneller Perspektive*. Berlin, Boston: De Gruyter. doi:10.1515/9783110536591.

Vasishth, Shravan & Andrew Gelman. 2021. How to embrace variation and accept uncertainty in linguistic and psycholinguistic data analysis. *Linguistics* 59(5). 1311–1342. doi:10.1515/ling-2019-0051.

Vihman, Virve-Anneli & George Walkden. 2021. Verb-second in spoken and written Estonian. *Glossa: a journal of general linguistics* 6(1). 1–23. doi:10.5334/gjgl.1404.

Vikner, Sten. 1995. *Verb movement and expletive subjects in the Germanic languages*. New York, Oxford: Oxford University Press. doi:10.1093/oso/9780195083934.001.0001.

Vikner, Sten. 2006. Object shift. In Martin Everaert & Henk Van Riemsdijk (eds.), *The Blackwell companion to syntax*, vol. III, 392–436. Malden, MA: Wiley. doi:10.1002/9780470996591.ch46.

Waldmann, Christian. 2008. *Input och output. Ordföljd i svenska barns huvudsatser och bisatser*. Lund: Lunds Universitet Dissertation.

Walkden, George. 2014. *Syntactic reconstruction and Proto-Germanic*. Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780198712299.001.0001.

Walkden, George. 2017a. Language contact and V3 in Germanic varieties new and old. *Journal of Comparative Germanic Linguistics* 20(1). 49–81. doi:10.1007/s10828-017-9084-2.

Walkden, George. 2017b. The actuation problem. In Adam Ledgeway & Ian Roberts (eds.), *The Cambridge handbook of historical syntax*, 403–424. Cambridge: Cambridge University Press. doi:10.1017/9781107279070.020.

Walkden, George. 2021a. Against mechanisms: Towards a minimal theory of change. *Journal of Historical Syntax* 5(33). 1–27. doi:10.18148/hs/2021.v5i32-39.141.

Walkden, George. 2021b. Do the wealthy stay healthy? Rich agreement and verb movement in early English. *Journal of Historical Syntax* 5(30). 1–28. doi:10.18148/hs/2021.v5i30.77.

Walkden, George. to appear. Scandinavians and verb-second in Northumbrian Old English. In Richard Dance, Sara M. Pons-Sanz & Brittany Schorn (eds.), *The legacy of Medieval Scandinavian encounters with England and the insular world*, 1–39. Turnhout: Brepols.

Wallan, Ashwag. 2018. *Evaluation of Arabic tests of sentence repetition and verbal short term memory for Saudi preschoolers*. London: City, University of London dissertation.

Wallenberg, Joel C., Rachael Bailes, Christine Cuskley & Anton Karl Ingason. 2021. Smooth signals and syntactic change. *Languages* 6(2). 1–20. doi:10.3390/languages6020060.

Watson, Stuart K., Judith M. Burkart, Steven J. Schapiro, Susan P. Lambeth, Jutta L. Mueller & Simon W. Townsend. 2020. Nonadjacent dependency processing in monkeys, apes, and humans. *Science Advances* 6(43). 1–9. doi:10.1126/sciadv.abb0725.

Weerman, Fred. 1989. *The V2 conspiracy. A synchronic and a diachronic analysis of verbal positions in Germanic languages*. Dordrecht, Providence, RI: Foris. doi:10.1515/9783110250442.

Westendorp, Maud. 2022. *The distribution of main and embedded structures. V2 and non-V2 orders in North Germanic*. Tromsø: UiT The Arctic University of Norway dissertation.

Westendorp, Maud & Björn Lundquist. 2022. Unstable verb placement and the structure of North Germanic main and embedded clauses. Manuscript.
Westergaard, Marit. 2003. Word order in wh-questions in a North Norwegian dialect: Some evidence from an acquisition study. *Nordic Journal of Linguistics* 26(1). 81–109. doi:10.1017/S0332586503001021.
Westergaard, Marit. 2008. Acquisition and change: On the robustness of the triggering experience for word order cues. *Lingua* 118(12). 1841–1863. doi:10.1016/j.lingua.2008.05.003.
Westergaard, Marit. 2009a. Microvariation as diachrony: A view from acquisition. *Journal of Comparative Germanic Linguistics* 12(1). 49–79. doi:10.1007/s10828-009-9025-9.
Westergaard, Marit. 2009b. *The acquisition of word order. Micro-cues, information structure, and economy*. Amsterdam, Philadelphia: John Benjamins. doi:10.1075/la.145.
Westergaard, Marit. 2009c. Usage-based vs. rule-based learning: The acquisition of word order in wh-questions in English and Norwegian. *Journal of Child Language* 36(5). 1023–1051. doi:10.1017/S0305000909009349.
Westergaard, Marit. 2009d. Word order in Old and Middle English: The role of information structure and first language acquisition. *Diachronica* 26(1). 65–102. doi:10.1075/dia.26.1.03wes.
Westergaard, Marit. 2010. Cue-based acquisition and information structure drift in diachronic language development. In Gisella Ferraresi & Rosemarie Lühr (eds.), *Diachronic studies on information structure. Language acquisition and change*, 87–116. Berlin, New York: De Gruyter. doi:10.1515/9783110227475.87.
Westergaard, Marit. 2011. Subject positions and information structure: The effect of frequency on acquisition and change. *Studia Linguistica* 65(3). 299–332. doi:10.1111/j.1467-9582.2011.01183.x.
Westergaard, Marit. 2014. Linguistic variation and micro-cues in first language acquisition. *Linguistic Variation* 14(1). 26–45. doi:10.1075/lv.14.1.02wes.
Westergaard, Marit. 2017. Gradience and gradualness vs abruptness. In Adam Ledgeway & Ian Roberts (eds.), *The Cambridge handbook of historical syntax*, 446–466. Cambridge: Cambridge University Press. doi:10.1017/9781107279070.022.
Westergaard, Marit. 2021a. Language acquisition, microcues, parameters, and morphosyntactic change. In Richard D. Janda, Brian D. Joseph & Barbara S. Vance (eds.), *The handbook of historical linguistics*, vol. 2, 357–374. Hoboken, NJ, Wiley Blackwell. doi:10.1002/9781118732168.ch17.
Westergaard, Marit. 2021b. Microvariation in multilingual situations: The importance of property-by-property acquisition. *Second Language Research* 37(3). 379–407. doi:10.1177/0267658319884116.
Westergaard, Marit, Terje Lohndal & Björn Lundquist. 2023. Variable V2 in Norwegian heritage language. An effect of crosslinguistic influence? *Linguistic Approaches to Bilingualism* 13(2). 133–162. doi:10.1075/lab.20076.wes.
Westergaard, Marit R. & Øystein A. Vangsnes. 2005. Wh-questions, V2, and the left periphery of three Norwegian dialect types. *Journal of Comparative Germanic Linguistics* 8(1). 117–158. doi:10.1007/s10828-004-0292-1.

White, James. 2014. Evidence for a learning bias against saltatory phonological alternations. *Cognition* 130(1). 96–115. doi:10.1016/j.cognition.2013.09.008.
Wickham, Hadley. 2016. *ggplot2. Elegant graphics for data analysis*. Cham: Springer. doi:10.1007/978-3-319-24277-4.
Wiese, Heike, Mehmet Tahir Öncü, Hans G. Müller & Eva Wittenberg. 2020. Verb third in spoken German. In Rebecca Woods & Sam Wolfe (eds.), *Rethinking verb second*, 682–699. Oxford: Oxford University Press. doi:10.1093/oso/9780198844303.003.0029.
Wiklund, Anna-Lena, Kristine Bentzen, Gunnar Hrafn Hrafnbjargarson & Þorbjörg Hróarsdóttir. 2009. On the distribution and illocution of V2 in Scandinavian that-clauses. *Lingua* 119(12). 1914–1938. doi:10.1016/j.lingua.2009.03.006.
Williams, John N., Peter Möbius & Choonkyong Kim. 2001. Native and non-native processing of English wh-questions: Parsing strategies and plausibility constraints. *Applied Psycholinguistics* 22(4). 509–540. doi:10.1017/S0142716401004027.
Willis, David. 1998. *Syntactic change in Welsh. A study on the loss of verb-second*. Oxford: Clarendon. doi:10.1093/oso/9780198237594.001.0001.
Willis, David. 2007. Specifier-to-head reanalyses in the complementizer domain: Evidence from Welsh. *Transactions of the Philological Society* 105(3). 432–480. doi:10.1111/j.1467-968X.2007.00194.x.
Winter, Werner. 1961. Relative Häufigkeit syntaktischer Erscheinungen als Mittel zur Abgrenzung von Stilarten. *Phonetica* 7(4). 193–216. doi:10.1159/000258110.
Wolfe, Sam. 2015a. Medieval Sardinian. New evidence for syntactic change from Latin to Romance. In Dag T. T. Haug (ed.), *Historical Linguistics 2013: Selected papers from the 21st International Conference on Historical Linguistics, Oslo, 5-9 August 2013*. 303–324. Amsterdam, Philadelphia: John Benjamins. doi:10.1075/cilt.334.16wol.
Wolfe, Sam. 2015b. The nature of Old Spanish verb second reconsidered. *Lingua* 164. 132–155. doi:10.1016/j.lingua.2015.06.007.
Wolfe, Sam. 2016. A comparative perspective on the evolution of Romance clausal structure. *Diachronica* 33(4). 461–502. doi:10.1075/dia.33.4.02wol.
Wolfe, Sam. 2018a. Occitan, verb second and the Medieval Romance word order debate. In Janine Berns, Haike Jacobs & Dominique Nouveau (eds.), *Selected papers from "Going Romance" 29, Nijmegen*, 315–336. Amsterdam, Philadelphia: John Benjamins. doi:10.1075/rllt.13.19wol.
Wolfe, Sam. 2018b. Probing the syntax of a problematic particle: Old French 'si' revisited. *Transactions of the Philological Society* 116(3). 332–362. doi:10.1111/1467-968X.12123.
Wolfe, Sam. 2018c. *Verb second in Medieval Romance*. Oxford: Oxford University Press. doi:10.1093/oso/9780198804673.001.0001.
Wolfe, Sam. 2019. Redefining the typology of V2 languages. The view from Medieval Romance and beyond. *Linguistic Variation* 19(1). 16–46. doi:10.1075/lv.15026.wol.
Wolfe, Sam. 2020. Rethinking Medieval Romance verb second. In Rebecca Woods & Sam Wolfe (eds.), *Rethinking verb second*, 348–367. Oxford: Oxford University Press. doi:10.1093/oso/9780198844303.003.0015.
Wolfe, Sam. 2021. Parametric stability and word order change: The case of Middle French. *Journal of Historical Syntax* 5(31). 1–44. doi:10.18148/hs/2021.v5i31.108.

Wöllstein, Angelika. 2010. *Topologisches Satzmodell*. Heidelberg: Winter.
Wonnacott, Elizabeth, Helen Brown & Kate Nation. 2017. Skewing the evidence: The effect of input structure on child and adult learning of lexically based patterns in an artificial language. *Journal of Memory and Language* 95. 36–48. doi:10.1016/j.jml.2017.01.005.
Yang, Charles. 2000. Internal and external forces in language change. *Language Variation and Change* 12(3). 231–250. doi:10.1017/S0954394500123014.
Yang, Charles. 2002. *Knowledge and learning in natural language*. New York: Oxford University Press. doi:10.1093/oso/9780199254149.001.0001.
Yang, Charles. 2010. Three factors in language variation. *Lingua* 120(5). 1160–1177. doi:10.1016/j.lingua.2008.09.015.
Yang, Charles. 2016. *The price of linguistic productivity*. Cambridge, MA: MIT Press. doi:10.7551/mitpress/9780262035323.001.0001.
Yin, Sora Heng & James White. 2018. Neutralization and homophony avoidance in phonological learning. *Cognition* 179. 89–101. doi:10.1016/j.cognition.2018.05.023.
Zimmerman, Frederick J., Jill Gilkerson, Jeffrey A. Richards, Dimitri A. Christakis, Dongxin Xu, Sharmistha Gray & Umit Yapanel. 2009. Teaching by listening: The importance of adult-child conversations to language development. *Pediatrics* 124(1). 342–349. doi:10.1542/peds.2008-2267.
Zimmermann, Michael. 2014. *Expletive and referential subject pronouns in Medieval French*. Berlin, Boston: De Gruyter. doi:10.1515/9783110367478.
Zipf, George Kingsley. 1935. *The psycho-biology of language: An introduction to dynamic philology*. Boston: Houghton Mifflin.
Zwart, C. Jan-Wouter. 1997. *Morphosyntax of verb movement. A minimalist approach to the syntax of Dutch*. Dordrecht, Boston, London: Kluwer. doi:10.1007/978-94-011-5880-0.
Þorgeirsson, Haukur. 2012. Late placement of the finite verb in Old Norse Fornyrðislag meter. *Journal of Germanic Linguistics* 24(3). 233–269. doi:10.1017/S1470542712000037.

APPENDIX

Appendix A

Appendix A contains all training and testing materials used in the experiments reported in Chapter 2 and Chapter 3. The production materials for testing were only used in experiment 2 (§2.4) and experiment 3 (§3.2).

Training materials

Set	Word order	Block	Sentence
1	S-V-M-O-A	A	The meteorologist checks automatically the map every morning.
1	O-V-S-M-A	B	The map checks the meteorologist automatically every morning.
1	A-V-S-M-O	C	Every morning checks the meteorologist automatically the map.
2	S-V-M-O-A	C	William devours usually a feast on Tuesday.
2	O-V-S-M-A	A	A feast devours William usually on Tuesday.
2	A-V-S-M-O	B	On Tuesday devours William usually a feast.
3	S-V-M-O-A	B	The critic tastes carefully a dish every Monday.
3	O-V-S-M-A	C	A dish tastes the critic carefully every Monday.
3	A-V-S-M-O	A	Every Monday tastes the critic carefully a dish.
4	S-V-M-O-A	A	Lillian reads always a story at bedtime.
4	O-V-S-M-A	B	A story reads Lillian always at bedtime.
4	A-V-S-M-O	C	At bedtime reads Lillian always a story.
5	S-V-M-O-A	C	The chef slices aggressively an onion upstairs.
5	O-V-S-M-A	A	An onion slices the chef aggressively upstairs.
5	A-V-S-M-O	B	Upstairs slices the chef aggressively an onion.
6	S-V-M-O-A	B	Edward discovers alone a lake in Cuba.
6	O-V-S-M-A	C	A lake discovers Edward alone in Cuba.
6	A-V-S-M-O	A	In Cuba discovers Edward alone a lake.
7	S-V-M-O-A	A	The musician plays proudly the guitar outside.
7	O-V-S-M-A	B	The guitar plays the musician proudly outside.
7	A-V-S-M-O	C	Outside plays the musician proudly the guitar.
8	S-V-M-O-A	C	Alice designs finally the poster in Lisbon.
8	O-V-S-M-A	A	The poster designs Alice finally in Lisbon.
8	A-V-S-M-O	B	In Lisbon designs Alice finally the poster.

Set	Word order	Block	Sentence
9	S-V-M-O-A	B	The lumberjack wields dramatically the ax every day.
9	O-V-S-M-A	C	The ax wields the lumberjack dramatically every day.
9	A-V-S-M-O	A	Every day wields the lumberjack dramatically the ax.
10	S-V-M-O-A	B	Adam cleans quickly the house during summer.
10	O-V-S-M-A	C	The house cleans Adam quickly during summer.
10	A-V-S-M-O	A	During summer cleans Adam quickly the house.
11	S-V-M-O-A	C	The toddler wears typically a dress on Sunday.
11	O-V-S-M-A	A	A dress wears the toddler typically on Sunday.
11	A-V-S-M-O	B	On Sunday wears the toddler typically a dress.
12	S-V-M-O-A	A	Emily unseals silently the bottle at midnight.
12	O-V-S-M-A	B	The bottle unseals Emily silently at midnight.
12	A-V-S-M-O	C	At midnight unseals Emily silently the bottle.
13	S-V-M-O-A	A	The miller grinds traditionally the grain in Nebraska.
13	O-V-S-M-A	B	The grain grinds the miller traditionally in Nebraska.
13	A-V-S-M-O	C	In Nebraska grinds the miller traditionally the grain.
14	S-V-M-O-A	C	Emmet rents sometimes an apartment in France.
14	O-V-S-M-A	A	An apartment rents Emmet sometimes in France.
14	A-V-S-M-O	B	In France rents Emmet sometimes an apartment.
15	S-V-M-O-A	B	The engineer develops efficiently a tool in Germany.
15	O-V-S-M-A	C	A tool develops the engineer efficiently in Germany.
15	A-V-S-M-O	A	In Germany develops the engineer efficiently a tool.
16	S-V-M-O-A	A	Georgia hides secretly the chocolate inside.
16	O-V-S-M-A	B	The chocolate hides Georgia secretly inside.
16	A-V-S-M-O	C	Inside hides Georgia secretly the chocolate.
17	S-V-M-O-A	C	The caregiver visits officially the zoo every Sunday.
17	O-V-S-M-A	A	The zoo visits the caregiver officially every Sunday.
17	A-V-S-M-O	B	Every Sunday visits the caregiver officially the zoo.
18	S-V-M-O-A	B	Connor collects happily a payment every Tuesday.
18	O-V-S-M-A	C	A payment collects Connor happily every Tuesday.
18	A-V-S-M-O	A	Every Tuesday collects Connor happily a payment.
19	S-V-M-O-A	A	The archeologist excavates actively the temple next month.
19	O-V-S-M-A	B	The temple excavates the archeologist actively next month.
19	A-V-S-M-O	C	Next month excavates the archeologist actively the temple.
20	S-V-M-O-A	C	Megan decorates nicely the tree in December.
20	O-V-S-M-A	A	The tree decorates Megan nicely in December.
20	A-V-S-M-O	B	In December decorates Megan nicely the tree.
21	S-V-M-O-A	B	The father sells hardly a motorcycle in Italy.
21	O-V-S-M-A	C	A motorcycle sells the father hardly in Italy.
21	A-V-S-M-O	A	In Italy sells the father hardly a motorcycle.

Set	Word order	Block	Sentence
22	S-V-M-O-A	A	Joe presents hopefully the product in Berlin.
22	O-V-S-M-A	B	The product presents Joe hopefully in Berlin.
22	A-V-S-M-O	C	In Berlin presents Joe hopefully the product.
23	S-V-M-O-A	C	The tourist buys quickly a villa in Bulgaria.
23	O-V-S-M-A	A	A villa buys the tourist quickly in Bulgaria.
23	A-V-S-M-O	B	In Bulgaria buys the tourist quickly a villa.
24	S-V-M-O-A	B	Tracy calculates precisely the equation in school.
24	O-V-S-M-A	C	The equation calculates Tracy precisely in school.
24	A-V-S-M-O	A	In school calculates Tracy precisely the equation.
25	S-V-M-O-A	A	The witch brews personally the potion since 2010.
25	O-V-S-M-A	B	The potion brews the witch personally since 2010.
25	A-V-S-M-O	C	Since 2010 brews the witch personally the potion.
26	S-V-M-O-A	C	Thomas causes unfortunately an accident on Friday.
26	O-V-S-M-A	A	An accident causes Thomas unfortunately on Friday.
26	A-V-S-M-O	B	On Friday causes Thomas unfortunately an accident.
27	S-V-M-O-A	B	The teacher corrects reluctantly the exam over night.
27	O-V-S-M-A	C	The exam corrects the teacher reluctantly over night.
27	A-V-S-M-O	A	Over night corrects the teacher reluctantly the exam.
28	S-V-M-O-A	A	Margaret earns currently a living in Switzerland.
28	O-V-S-M-A	B	A living earns Margaret currently in Switzerland.
28	A-V-S-M-O	C	In Switzerland earns Margaret currently a living.
29	S-V-M-O-A	C	The author revises eventually a novel in Boston.
29	O-V-S-M-A	A	A novel revises the author eventually in Boston.
29	A-V-S-M-O	B	In Boston revises the author eventually a novel.
30	S-V-M-O-A	B	Ethan repairs fortunately the toilet downtown.
30	O-V-S-M-A	C	The toilet repairs Ethan fortunately downtown.
30	A-V-S-M-O	A	Downtown repairs Ethan fortunately the toilet.

Testing materials

Production

Set type	Subject	Verb	Marker	Direct Object	Adjunct/Indirect object
simple A	the driver	delivers	grumpily	the food	this afternoon
simple A	Sophia	sings	loudly	a carol	on Christmas
simple A	the general	declares	hastily	victory	in Waterloo
simple A	Sam	drinks	delightedly	espresso	in Milano
complex A	the secretary	conceals	ruthlessly	the evidence	during the conflict
complex A	Brianna	refutes	energetically	the rumor	in late January
complex A	the invalid	takes	usually	the medicine	in the kitchen
complex A	Jayden	sweeps	halfheartedly	the floor	in the bathroom
IO	the waiter	passes	awkwardly	the saltshaker	to the guest
IO	Sarah	tells	laughingly	a joke	to the child
IO	the investigator	submits	contentedly	the report	to the prosecutor
IO	Charles	suggests	cheekily	a whiskey	to the friend

Judgement

V-pos	Initial	Word order	Sentence
V2	DO	O-V-S-M-A	A conference holds the superintendent officially in November.
V2	DO	O-V-S-M-A	The statement approves Harper hopefully this week.
V2	DO	O-V-S-M-A	The material understands the student poorly at home.
V2	DO	O-V-S-M-A	A message retypes Hailey obsessively on whatsapp.
V2	simple A	A-V-S-M-A	Tonight steals the villain secretly a diamond.
V2	simple A	A-V-S-M-A	While cooking obeys Evelyn hardly the recipe.
V2	simple A	A-V-S-M-A	In Atlanta approaches the pilot slowly the runway.
V2	simple A	A-V-S-M-A	At bingo selects Wesley randomly a number.
V2	IO	IO-V-S-M-DO	To his son sends John personally a book.
V2	IO	IO-V-S-M-DO	To her daughter gives Mary always the check.
V2	IO	IO-V-S-M-DO	To the congregation shows the priest silently the candle.
V2	IO	IO-V-S-M-DO	To the coach throws the player gently the ball.
V2	complex A	AC-V-S-M-O	In late April regrets the politician openly his misconduct.
V2	complex A	AC-V-S-M-O	On this weekend admits the actor publicly the addiction.
V2	complex A	AC-V-S-M-O	In conservative Utah repeats the minister properly the oath.
V2	complex A	AC-V-S-M-O	On ad-free TV promotes the host surprisingly a product.
V3	DO	O-S-V-M-A	The magazine the editor publishes traditionally every Thursday.
V3	DO	O-S-V-M-A	Breakfast Larry eats normally before sunrise.
V3	DO	O-S-V-M-A	The sale the administrator regulates voluntarily in Austria.
V3	DO	O-S-V-M-A	The graduation Emma celebrates adequately in Florida.
V3	simple A	A-S-V-M-O	After Christmas the agent investigates actively the murder.
V3	simple A	A-S-V-M-O	Before sunset Jack folds oddly the T-shirt.
V3	simple A	A-S-V-M-O	In Reno the magician bends successfully the spoon.
V3	simple A	A-S-V-M-O	In Springfield Rachel renews constantly the policy.

V-pos	Initial	Word order	Sentence
V3	IO	IO-S-V-M-DO	To the doctor the patient describes precisely the pain.
V3	IO	IO-S-V-M-DO	To the shop George returns safely the computer.
V3	IO	IO-S-V-M-DO	To a beggar the tailor donates repeatedly a coat.
V3	IO	IO-S-V-M-DO	To the co-worker Sylvia mentions privately the secret.
V3	complex A	IO-S-V-M-DO	At the moment the referee verifies briefly the decision.
V3	complex A	IO-S-V-M-DO	This Thursday afternoon the artist finishes already a painting.
V3	complex A	IO-S-V-M-DO	In every corner the boy sees apparently peril.
V3	complex A	IO-S-V-M-DO	In the workshop the carpenter saws rarely a plank.

Appendix B

In Chapter 3, the results of a Monte Carlo simulation were briefly discussed in footnote 16. Here, the simulation will be discussed in more detail. The discussion is based on a subsection that was included in an earlier draft of the paper in §3.2 but later dropped for reasons of space. Consequently, the same authorship statement as in Chapter 3 applies.

The results of the large-scale corpus study outlined in §3.2 have demonstrated that the frequency distribution of clause-initial constituents in German is skewed. Subjects are the most frequent clause-initial element, followed by adjuncts. Direct objects and other constituents types are rarely found in initial position (cf. Figure 3.8). However, it is worth considering what might drive this frequency skew, and what exactly learners conclude from it. In particular, at least some of the skew may result from the probability with which a given element is present in a sentence in the first place. If one type of constituent has a higher baseline frequency of occurrence than another, it has more opportunity to be fronted. For example, subjects are in principle more likely to occur in sentences than objects because subjects are expressed in both intransitive and transitive constructions.[1] In a sense it is therefore unsurprising that subjects are the dominant clause-initial element. More generally, it is possible that different constituent types are actually equally likely to appear in clause-initial position once their likelihood of appearing in a sentence is incorporated; in other words, the conditional probability of different types may be identical. There is good reason to believe that learners are sensitive to conditional probabilities, and can track and use them to learn about linguistic structure (Saffran, Aslin & Newport 1996, Aslin, Saffran & Newport 1998). If learners take into account conditional probabilities this might influence how they perceive the distribution of clause-initial elements in a language like German. Therefore it is worth determining whether the skew persists once this is taken into account.

In order to determine the conditional probabilities of different constituent types appearing clause-initially, I ran a Monte Carlo simulation on the data reported in §3.2.[2]

[1] This glosses over the fact that topic drop (ia) and null subjects (ib) are permissible in certain contexts in German (Cardinaletti 1990, Trutkowski 2016).

(i) a. _ **hab** ich gegessen. GERMAN
_ have I eaten
'I ate it'
b. ∅ **esse** heute nur Maultaschen.
pro eat today only Swabian ravioli
'I only eat Swabian ravioli today.'

Despite the existence of constructions like (i), my argument still holds as they are not equally acceptable in all situations (Schäfer 2021).

[2] The conditional probability can also be calculated mathematically using (i). However, the total number of adjuncts in the corpora exceeds the number of clauses and, as a result, the conditional probability would be greater than 1. The simulation provides us with an elegant solution to circumvent this issue.

(i) $p(A|B) = \frac{p(A \cap B)}{p(B)}$.

Monte Carlo methods involve repeated random sampling which, in the present case, means sampling of clause-initial constituents from the set of frontable constituents in a clause. This process is repeated multiple times such that a distribution over distributions can be formed for each constituent type. The observed counts are then compared to the distribution of simulated counts. If the probability of being fronted is similar for each constituent, the observed counts for each should lie within the simulated distribution. If, on the other hand, the conditional probabilities diverge, the observed counts should lie outside the simulated distribution.

In a first step, all constituents that can appear clause-initially were extracted for each V2 clause. The constituent types and clauses were identical to the ones reported in §3.2. Next, one constituent was randomly chosen as clause-initial element for each sentence using a python script.[3] After each round the distribution of each constituent type was calculated by corpus. I repeated this process 10,000 times and compared the distribution over distributions to the observed counts afterwards. Again, if all observed counts lie outwith the distribution of the simulated counts, I can conclude that the skew noted in the corpus persists even when conditional probability of different categories are approximately equal.

A comparison of the observed counts to the simulated frequency distribution revealed a consistent pattern in that each of the observed frequencies (shown as red vertical lines) lies outside the simulated distribution (visualised in green). The direction of the results, however, varied between the different constituent types. Non-clausal subjects (Figure A.1), expletives (Figure A.2) and adverbial clauses (Figure A.3) occur more frequently in initial position than expected. For adjuncts, a split picture emerged (Figure A.4). Adjuncts occurred in initial position more frequently than expected in the wiki and europarl corpora but the opposite was the case in the speeches corpus. All remaining types such as direct objects (Figure A.5) and indirect objects (Figure A.6) occurred less frequently than expected across the three corpora. Hence, we can conclude that even when conditional probabilities are considered, the distribution of different constituent types in clause-initial position is still skewed. This result, in combination with the results obtained in §3.2, suggests that German speakers maintain a V2 grammar despite the skewed frequency of initial constituents.

[3]Importantly, non-initial conjuncts in coordinated structures were treated differently if a coordinator (*und* 'and', *oder* 'or') immediately preceded the finite verb (i.e. TP-coordination) as no constituent can be preposed in these contexts. All constituents were counted as postverbal in theses cases.

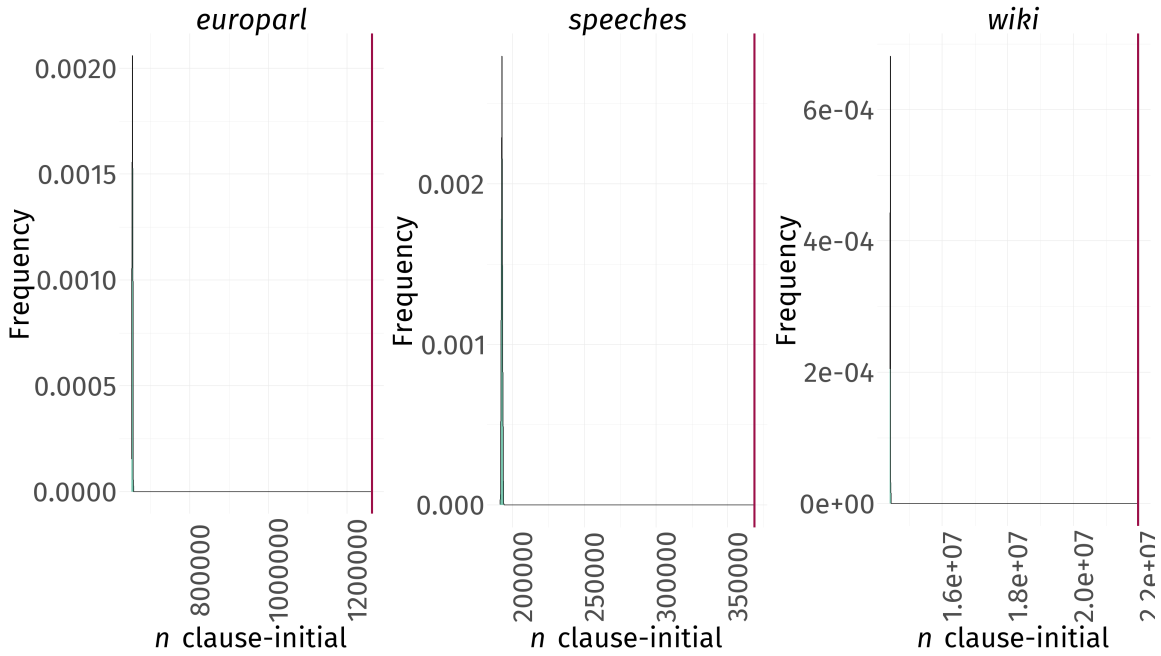


Figure A.1: Comparison of simulated and observed frequencies of clause-initial non-clausal subjects.

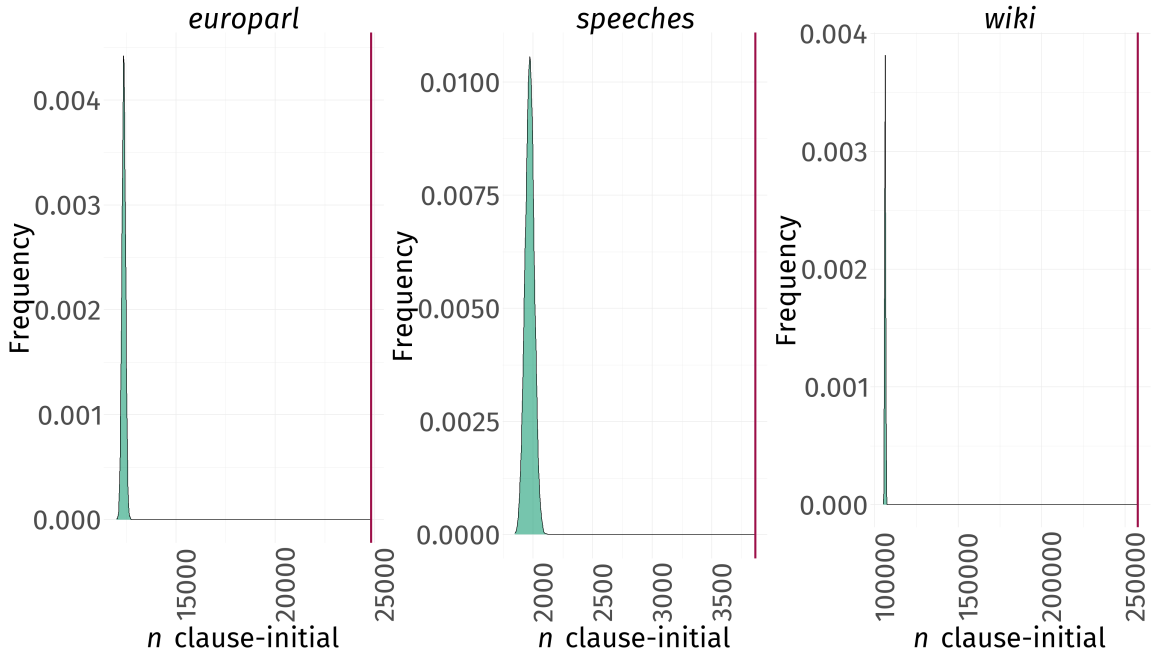


Figure A.2: Comparison of simulated and observed frequencies of clause-initial expletives.

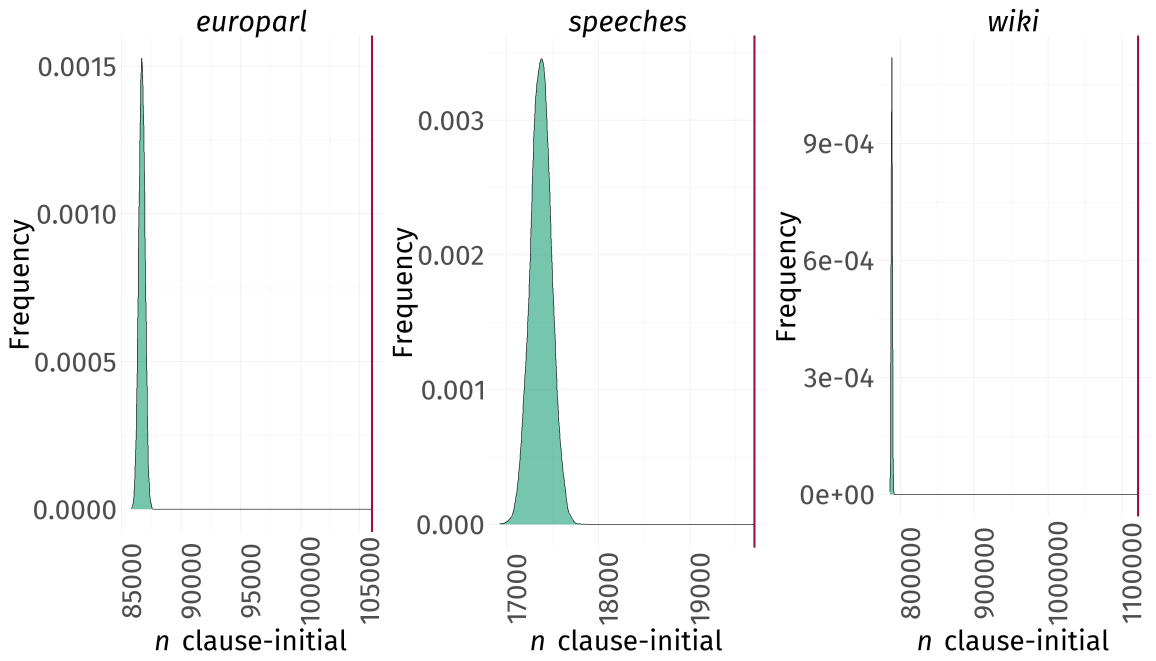


Figure A.3: Comparison of simulated and observed frequencies of clause-initial adverbial clauses.

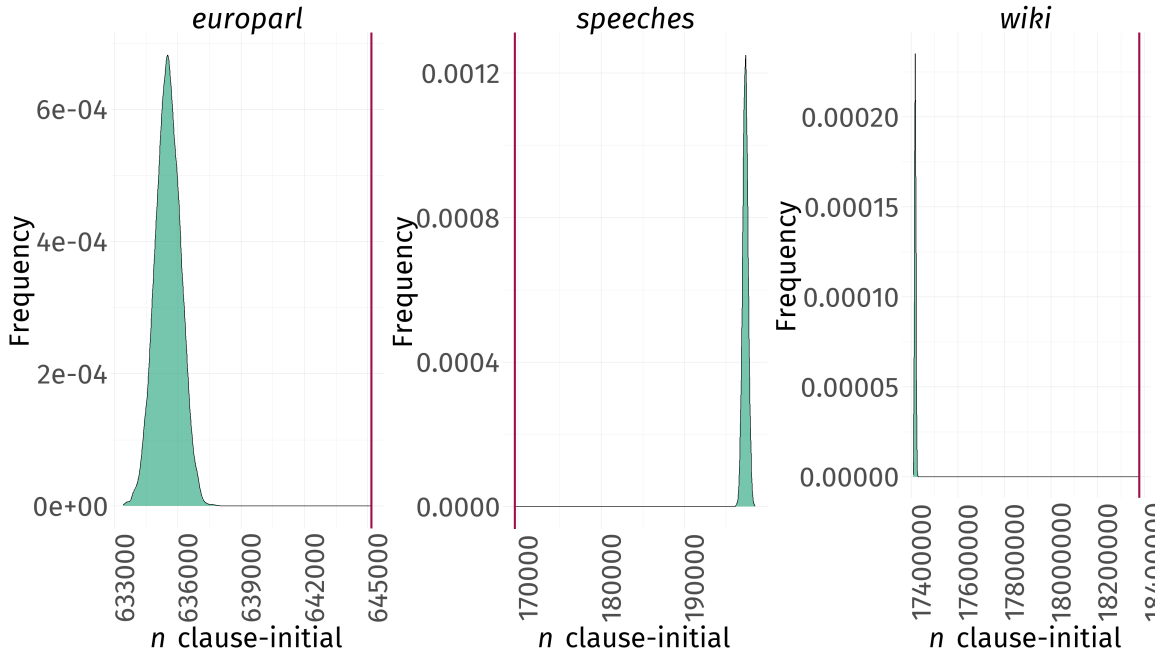


Figure A.4: Comparison of simulated and observed frequencies of clause-initial adjuncts.

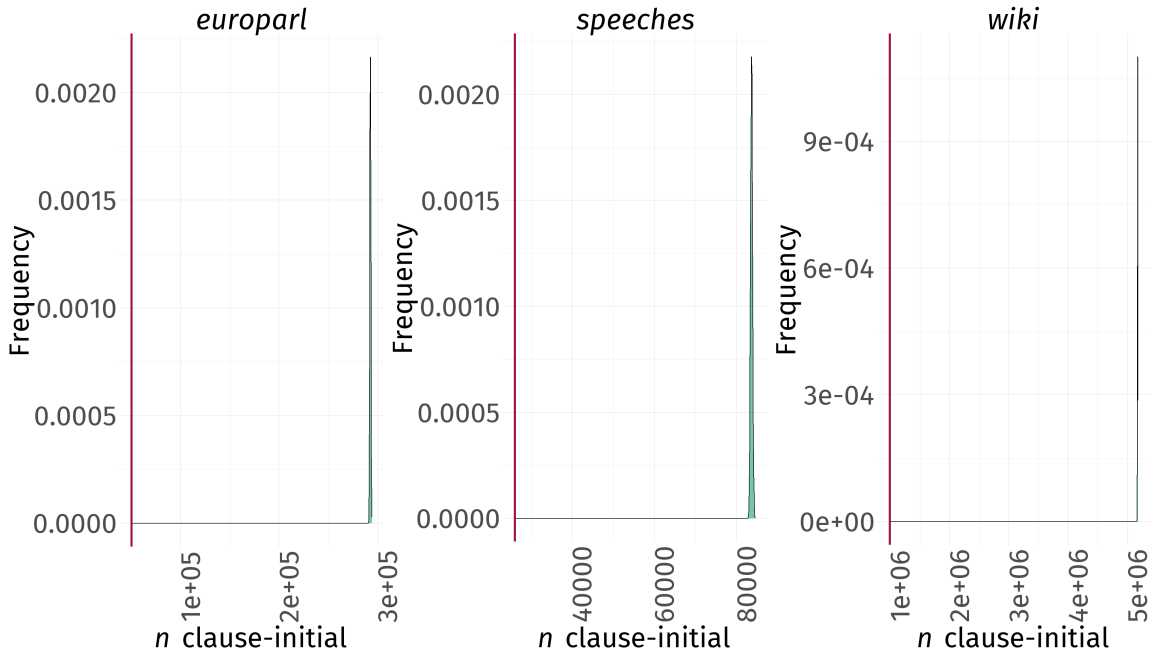


Figure A.5: Comparison of simulated and observed frequencies of clause-initial direct objects.

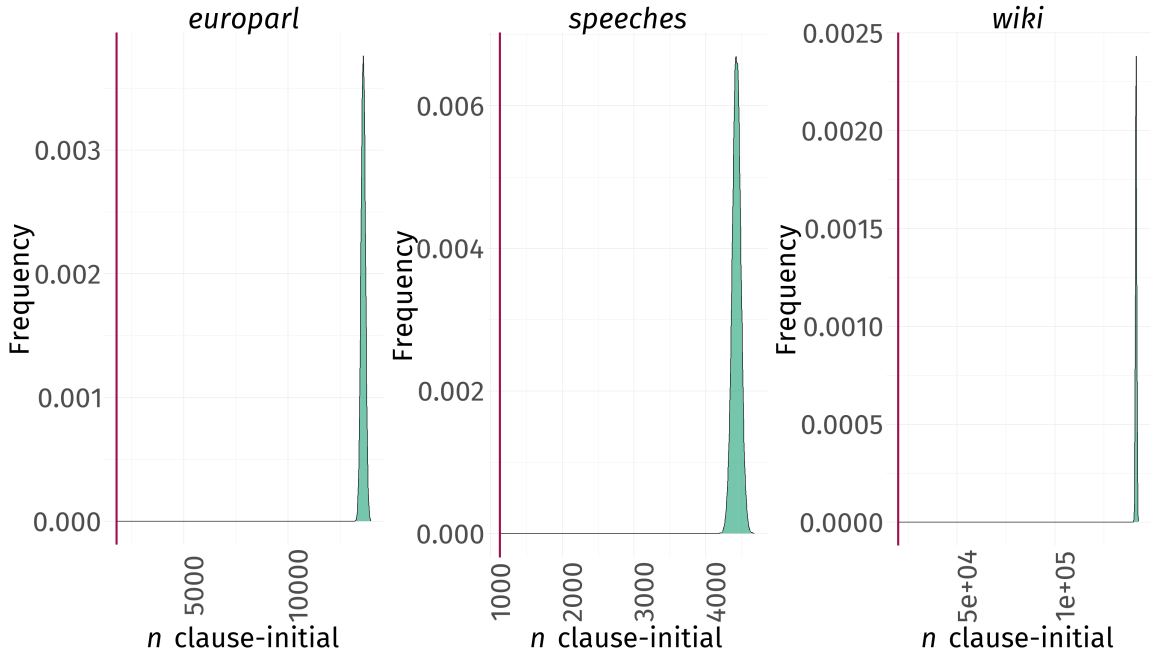


Figure A.6: Comparison of simulated and observed frequencies of clause-initial indirect objects.

Appendix C

Appendix C contains the training and testing materials of experiment 5 (§4.4). To increase the readability of the stimuli sentences, their English literal translations are provided.

Lexicon

Noun (animate)	Noun (inanimate)	PP	Verb
angel	bag	in the mountains	cook
chef	ball	on the beach	cut
boxer	book	on the boat	eat
construction worker	bowl	on the bridge	hammer
cyclist	box	in the cave	kiss
devil	bread	in the city	paint
diver	carrot	in the desert	photograph
doctor	chair	under the Eifel tower	see
captain	chocolate	on the farm	touch
ballerina	cushion	in the kitchen	wipe
pirate	glasses	at night	give
priest	hat	in Pisa	throw
police officer	keyboard	under the rainbow	
queen	leek	in the snow	
sheriff	mug	in the sunset	
student	plate	in the thunderstorm	
wizard	potato	on the train	
scientist	shoe	in the vineyard	

Training materials

Set	Word order	Sentence
1	S-V-DO-A	The angel paints the hat in the mountains
1	DO-V-S-A	The hat paints the angel in the mountains
1	A-V-S-DO	In the mountains paints the angle the hat
2	S-V-DO-A	The boxer paints the chair on the boat
2	DO-V-S-A	The chair paints the boxer on the boat
2	A-V-S-DO	On the boat paints the boxer the chair
3	S-V-DO-A	The cyclist eats the bread in the cave
3	DO-V-S-A	The bread eats the cyclist in the cave
3	A-V-S-DO	In the cave eats the cyclist the bread
4	S-V-DO-A	The diver eats the chocolate in the desert
4	DO-V-S-A	The chocolate eats the diver in the desert
4	A-V-S-DO	In the desert eats the diver the chocolate
5	S-V-DO-A	The captain kisses the book on the farm
5	DO-V-S-A	The book kisses the captain on the farm
5	A-V-S-DO	On the farm kisses the captain the book
6	S-V-DO-A	The pirate kisses the plate at night
6	DO-V-S-A	The plate kisses the pirate at night
6	A-V-S-DO	At night kisses the pirate the plate
7	S-V-DO-A	The police officer hammers the ball under the rainbow
7	DO-V-S-A	The ball hammers the police officer under the rainbow
7	A-V-S-DO	Under the rainbow hammers the police officer the ball
8	S-V-DO-A	The scientist hammers the mug in the snow
8	DO-V-S-A	The mug hammers the scientist in the snow
8	A-V-S-DO	In the snow hammers the scientist the mug
9	S-V-DO-A	The student cuts the leek in the thunderstorm
9	DO-V-S-A	The leek cuts the student in the thunderstorm
9	A-V-S-DO	In the thunderstorm cuts the student the leek
10	S-V-DO-A	The ballerina cuts the bag in the vineyard
10	DO-V-S-A	The bag cuts the ballerina in the vineyard
10	A-V-S-DO	In the vineyard cuts the ballerina the bag
11	S-V-DO-A	The chef touches the cushion on the beach
11	DO-V-S-A	The cushion touches the chef on the beach
11	A-V-S-DO	On the beach touches the chef the cushion
12	S-V-DO-A	The doctor touches the box under the Eiffel tower
12	DO-V-S-A	The box touches the doctor under the Eiffel tower
12	A-V-S-DO	Under the Eiffel tower touches the doctor the box

Testing materials

Production

Set type	Lexical novelty	Subject	Verb	Direct object	Adjunct/Indirect object
familiar	familiar	the scientist	wipes	the book	in the mountains
familiar	familiar	the student	sees	the hat	on the boat
familiar	familiar	the ballerina	photographs	the chair	in the cave
familiar	familiar	the chef	cooks	the leek	in the desert
familiar	novel	the construction worker	kiss	the bowl	on the bridge
familiar	novel	the devil	paints	the keyboard	in the city
familiar	novel	the priest	cuts	the carrot	in the kitchen
familiar	novel	the queen	touches	the glasses	in Pisa
novel	familiar	the diver	gives	the bread	to the doctor
novel	familiar	the captain	throws	the chocolate	to the cyclist
novel	familiar	the pirate	gives	the ball	to the boxer
novel	familiar	the police officer	throws	the plate	the the angel
novel	novel	the sheriff	gives	the potato	to the construction worker
novel	novel	the wizard	throws	the shoe	to the devil
novel	novel	the priest	gives	the bowl	to the wizard
novel	novel	the queen	throws	the keyboard	the the sheriff

Judgement

Order	Initial	Word order	Sentence
V2	S	S-V-DO-A	The construction worker touches the keyboard on the farm
V2	S	S-V-DO-A	The devil wipes the mug in the sunset
V2	S	S-V-DO-A	The police officer sees the glasses in the train
V2	DO	DO-V-S-A	The priest kisses the potato under the rainbow
V2	DO	DO-V-S-A	The box photographs the queen on the bridge
V2	DO	DO-V-S-A	The carrot cooks the doctor in the city
V2	A	A-V-S-DO	At night paints the sheriff the shoe
V2	A	A-V-S-DO	In the kitchen cooks the wizard the leek
V2	A	A-V-S-DO	In Pisa wipes the student the bowl
V2	IO	IO-V-S-DO	To the angel gives the devil the bowl
V2	IO	IO-V-S-DO	To the boxer throws the priest the hat
V2	IO	IO-V-S-DO	To the cyclist gives the student the chair
V2	IO	IO-V-S-DO	To the diver throws the ballerina the show
V2	IO	IO-V-S-DO	To the wizard gives the construction worker the glasses
V2	IO	IO-V-S-DO	To the devil throws the sheriff the bread
V2	IO	IO-V-S-DO	To the queen gives the chef the chocolate
V2	IO	IO-V-S-DO	To the sheriff throws the scientist the plate
V2	IO	IO-V-S-DO	To the construction worker gives the angel the book
V3	S	S-A-V-DO	The construction worker in the thunderstorm hammers the shoe
V3	S	S-A-V-DO	The devil on the bridge sees the bag
V3	S	S-A-V-DO	The chef in Pisa photographs the keyboard
V3	DO	DO-S-V-A	The potato the queen cuts in the snow
V3	DO	DO-S-V-A	The cushion the priest wipes on the train
V3	DO	DO-S-V-A	The glasses the scientist photographs in the city
V3	A	A-S-V-DO	In the vineyard the wizard eats the carrot
V3	A	A-S-V-DO	In the sunset the sheriff cooks chocolate
V3	A	A-S-V-DO	In the kitchen the pirate sees the bowl
V3	IO	IO-S-V-DO	To the captain the priest throws the keyboard
V3	IO	IO-S-V-DO	To the pirate the wizard gives the mug
V3	IO	IO-S-V-DO	To the police officer the boxer throws the leek
V3	IO	IO-S-V-DO	To the doctor the cyclist gives the potato
V3	IO	IO-S-V-DO	To the devil the construction worker throws the carrot
V3	IO	IO-S-V-DO	To the priest the queen gives the ball
V3	IO	IO-S-V-DO	To the queen the diver throws the bag
V3	IO	IO-S-V-DO	To the sheriff the captain gives the cushion
V3	IO	IO-S-V-DO	To the wizard the ballerina throws the box