MASTERARBEIT | MASTER'S THESIS

Titel | Title

Little Kids, Big Constructions: Slot-and-Frame Patterns and Spatial Language Acquisition in Early Multiword Speech

verfasst von | submitted by

Fanny Rössler BA

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Master of Arts (MA)

Wien | Vienna, 2025

Studienkennzahl lt. Studienblatt | Degree programme code as it appears on the student record sheet:	UA 066 812
Studienrichtung lt. Studienblatt | Degree programme as it appears on the student record sheet:	Masterstudium English Language and Linguistics
Betreut von | Supervisor:	Univ.-Prof. Dr. Mathilde Eveline Keizer

Table of Contents

List of Figures

List of Tables

Abbreviations and Conventions

ATT	ATTRIBUTE or ATTRIBUTE slot
CDS	child-directed speech
CHAT	*Codes for Human Analysis of Transcripts*
CHILDES	*CHIld Language Data Exchange System*
CLAN	*Child Language ANalysis*
CS	child speech
DIR	DIRECTION or DIRECTION slot
LM	landmark
LOC	LOCATION or LOCATION slot
MLU	mean length of utterance
PRO	PROCESS or PROCESS slot
REF	REFERENT or REFERENT slot
TIME	TIME or TIME slot
TR	trajector

Note:

Children's ages are indicated following the structure years;months, e.g. 2;6 means 2 years and 6 months.

The speech samples were transcribed using the CHAT notation system; [*] refers to errors and [?] to best guesses, for example (for details on the transcription conventions see MacWhinney, 2024).

Linguistic examples from the datasets are given in cursive and cited following the structure (Name of file, utterance number), e.g. (Eleanor 020600a, 195) relates to utterance 195 in Eleanor's test corpus named 020600a. Lieven and Goh's (n.d.-a, n.d.-b) original recordings and transcripts can be accessed via these links: https://childes.talkbank.org/access/Eng-UK/MPI-EVA-Manchester.html and https://childes.talkbank.org/access/Eng-UK/Thomas.html.

1. Introduction

Human language is distinct from communication by other animal species because it is symbolic and grammatical. According to the usage-based theory of language acquisition, however, children do not know how to use linguistic symbols in a patterned manner from birth (Tomasello, 2003). To interact with interlocutors and convey messages, they must learn the constructions of their target language, that is, how form maps onto meaning and vice versa (Goldberg, 2019). With domain-general cognitive skills like intention-reading and categorisation in place, learners produce language in the shape of rote-learned repetitions or utterances with minimal variations (Tomasello, 2000):

> When young children have something they want to say, they sometimes have a set expression readily available and so they simply retrieve that expression from their stored linguistic experience. When they have no set expression readily available, they retrieve linguistic schemas and items that they have previously mastered (either in their own production or in their comprehension of other speakers) and then 'cut and paste' them together as necessary for the communicative situation at hand. (Tomasello, 2000, p. 77)

Bannard and Lieven (2009), too, argue that children repeat and reuse numerous high-frequency chunks. Research on multiword strings has further emphasised that both child language and child-directed speech are heavily influenced by frames, which facilitate the acquisition of grammatical categories and larger patterns (Theakston & Lieven, 2017; Arnon, 2021).

At the same time, spatial language studies have revealed that children learn spatial terms early (e.g. Tomasello, 1987), along with the corresponding concepts (e.g. Bowerman & Choi, 2003), and use various strategies to talk about their environment (Ferrara et al., 2011; Yang & Pan, 2021). Yet very little is currently known about the schematisation of such spatial constructions.

The method applied in this thesis is the traceback approach, an operationalisation of the view that early language learning is slow, initially less creative than one might think and strongly related to precedents produced by the child or caregiver (Dąbrowska & Lieven, 2005). Goldberg (2009, p. 217) writes that this approach is "very exciting and may well ultimately lead to clear constraints on the production of novel utterances". There have been numerous implementations of the model (e.g. Lieven et al., 2009; Kol et al., 2013; Koch, 2019), although they all provide slightly different definitions. I attempt to relate three children's target utterances at 2;6 to what they have previously heard or produced. The Eleanor-, Fraser- and Thomas-corpus analysed in this thesis are subcorpora designed on the basis of the MPI-EVA-Manchester Corpus (Lieven & Goh, n.d.-a) and the Thomas Corpus (Lieven & Goh, n.d.-b); both are accessible via the CHILDES talkbank (MacWhinney & Snow, 1985). The strings to be traced back are restricted to those pertaining to linguistic descriptions of space in order to better understand how spatial utterances are acquired and abstracted across. I use the analysis tool CLAN (MacWhinney, 2025) to perform the tracebacks and interpret the results qualitatively afterwards.

Behrens (2011, p. 252-253) posits that the field of developmental linguistics has historically dealt with the *what* ('was'), *how* ('wie') and *why* ('warum') of child language acquisition. In the present study, I ask how much of the children's linguistic output can be explained using fixed strings, lexically specific schemas with open slots and the operations SUBSTITUTE, SUPERIMPOSE and ADD. In accordance with assumptions widely shared by usage-based and constructionist researchers, I expected most productions to be either frozen, i.e. preconstructed formulae, or schematic, i.e. slot-and-frame patterns. I also seek to determine the operations used and slots represented in the data, paying special attention to LOCATION and DIRECTION slots, which are severely understudied. Like Behrens (2011), I thus intend to address the *what*, *how* and *why* of spatial construction learning. This thesis will generate fresh insight into spatial multiword units during the early multiword stage of language development. It will also make an original contribution by characterising the link between acquiring constructions and acquiring spatial language.

After this introduction, the remaining thesis is organised as follows. It begins by giving a concise overview of the theoretical approaches relevant to the investigation of early spatial language constructions; the main points addressed are (a) child language acquisition, (b) construction grammar and (c) the language of space. Section 3 reviews previous studies on multiword sequences and spatial terms, while part 4 presents the traceback method as well as my own procedure and datasets. My analysis and key findings are laid out in section 5, examining spatial language repertoires, successfully reconstructed utterances and failed derivations. I go on to then relate those outcomes to the research objectives and aforementioned previous research. Finally, the conclusion provides a summary of the results and significance of this small exploratory study and discusses its limitations and implications for future research.

2. Theoretical Background

In the background section, I start by discussing the three areas of interest to this study, language acquisition, construction grammar and spatial language, individually before they are joined in the main part the thesis.

2.1. Baby Steps: Child Language Acquisition

Acquiring a language is often perceived as a uniform and straightforward task when, in fact, learners are required to master a number of different skills. Children notice the speech sounds of their target language, before they learn to produce and combine them; they form words, which, in turn, must be linked to meanings. Linguistic material is, in addition, productively stringed together into sentences until the learner can successfully express complex ideas and converse with interlocutors (Rowland, 2014, p. 2). Scholars continue to marvel at the speed and seeming effortlessness with which children pick up their first language. In Rowland's (2014, p. 1) words:

> It has become a trite phrase, but it is no less true for that: human children acquire the most complex communication system known to man before they learn to tie their shoelaces. How do children do this? How do children acquire language? The short answer is we do not know. We are not even close to an answer.

Historically, this question has been addressed frequently; some sources report that pre-scientific investigations, albeit scarce and disputable, have been conducted from as early as 700 B.C. (Campbell, 2006, p. 391). Interest in how children learn to speak gathered pace in the 1700s and 1800s as the number of published diary studies, observational work and lexical inventories grew (Clark, 2019, pp. 2-3). More systematic research exploring the factors associated with language development dates back to the beginning of the 20th century, encompassing, above all, descriptive, normative and theoretical discussions. With more information on acquisition available, the discipline has since expanded to cover a larger variety of topics and methods; it has, for instance, embraced the use of computational models, considered neurological bases and formulated practical recommendations for language learning (Shatz, 2007, pp. 1-3).

2.1.1. Models of Language Acquisition

In order for a child to acquire a language, certain prerequisites must be fulfilled: a brain wired for language and the right environment (Rowland, 2014, p. 13). Karmiloff-Smith (1998, p. 389) opines that today “[a]ll scientists […] —from the staunchest Chomskyan nativist to the most domain-general empiricist—agree that development involves contributions from both genes and environment”. While most researchers concede that, their relative importance for linguistic outcomes is still hotly debated. Previous research on child language acquisition largely falls into two schools of thought: the nativist, generativist, Universal Grammar (UG) approach and the constructivist, emergentist, socio-pragmatic, functionalist, usage-based approach (Ambridge & Lieven, 2011).[1]

The generativist[2] research tradition was popularised by Chomsky in the 1960s and asserts that humans are equipped with a blueprint for grammar by birth (Clark, 2019, p. 7). Chomsky’s (1957) generative grammar caused a shift in the study of language by moving away from usage and semantics to prioritise competence and syntax. He argues that there is an innate, rule-governed system in place that allows humans to produce and understand infinite original sentences. As children become adept in a language quickly at a young age and without much intervention, it must be an inborn trait, a language instinct (Pinker, 1994). The poverty of stimulus argument further states that infants are not exposed to sufficient input to explain the plethora of grammatical manoeuvres. Similarly, they do not receive enough negative feedback to recover from mistakes and learn a grammar from scratch (Saxton, 1996). This leads generativists to believe that categories and rules must be specified in advance of linguistic experience (Chomsky, 1965, 1980; Crain, 1991).

Chomsky (1965, p. 34) therefore assumes the existence of a ‘language-acquisition device’ (LAD), intrinsic structures that are activated upon confrontation with a language: “To acquire language, a child must devise a hypothesis compatible with presented data—he must select from the store of potential grammars a specific one that is appropriate to the data available to him”. Such an inbuilt faculty for languages would need to be constrained yet flexible to account for the manifoldness of languages; thus, universal grammar (UG) was coined, which suggests that the same parameters underlie all languages (Chomsky, 1980, pp. 8-9). Crain (1991, p. 597)

[1] Kidd and Garcia (2022) find that child language research over the last 50 years is heavily skewed towards the Global North; as a result, most assumptions about acquisition are derived from just 2 percent of languages. The longstanding dominance of English and other widely studied Indo-European languages is reflected in theories of language learning, as the short overview hereafter illustrates.

[2] Most generativist work is necessarily nativist in nature as it posits that infants are born with linguistic knowledge. Not all nativist approaches, however, subscribe to the generativist assumption that grammar is innate but rather other language areas. Universal Grammar approaches marry both generativist and nativist ideas (Ambridge & Lieven, 2011, p. 2). I will subsume these approaches under the term generativism or generativist theories.

describes the acquisitional scenario schematically: “Input (PLD) → LAD → Final State”. Primary linguistic data (PLD) is mapped into the grammatical system of knowledge to reach the final state of adult grammar (Crain, 1991, pp. 597-598). In all, generativist theories present biology as the engine of development.

This generativist approach to language acquisition, although revolutionary, was nonetheless not accepted throughout the linguistic community; several linguists took issue with its theoretical assumptions of language universals, the appreciation of abstract innate principles over environment and the proposal of a specialised language module (e.g. Elman et al., 1996; Tomasello, 2003; Hoff, 2006). The usage-based paradigm, closely allied with cognitive linguistics and construction grammar (see section 2.2.), developed as a response and argues that language is learned from language use (Koch, 2019, p. 13). Ambridge et al. (2014) evaluate whether a universal grammar would actually help the learner in acquiring a language, concluding that other mechanisms and cues are decisive. They further identify faults in the generativist argumentation related to linking input to innate labels, data coverage and redundancy (Ambridge et al., 2014, pp. 81-82). Ibbotson (2022), moreover, writes that due to the linguistic diversity of the world’s 7000-plus languages it is difficult to define which linguistic content could be written into a universal grammar. Some biases facilitating learning one language might disadvantage the next; instead, he makes a case for universal psychology.

While the human brain is arguably configured for human language, usage-based developmentalists reject the idea that children are endowed with a priori knowledge of categories, principles and parameters (Elman et al., 1994, p. 46). Hoff (2006) gives a comprehensive review of social factors influencing language acquisition, like quality and quantity of parental input, socio-economic and ethnic background, birth order, schooling or interactions with peers. She finds that all contexts that provide children with the opportunity to hear analysable speech and have communicative experiences nurture the acquisition process; knowledge structures are seen as triggered by input and not instantaneous (Hoff, 2006, p. 55).

According to Tomasello (2003) and Christiansen and Chater (2008), grammar emerges from domain-general processes as opposed to an autonomous module for language, a “linguistic toolkit” with innately specified algebraic rules (Jackendoff, 2002, p. 79). Tomasello (2003, pp. 3-4) describes how the acquisition process is guided by learning mechanisms that allow children to develop skills of intention-reading and pattern-finding. Bybee (2010, pp. 11-12) also mentions categorisation, chunking, rich memory storage, analogy and cross-modal association as general-purpose cognitive skills underlying language development. In short, humans "derive language from non-language" (Lindblom et al., 1984, p. 187).

Attempts have been made to reconcile the two approaches, arguing that some linguistic learning as well as some innate knowledge is required. While this statement has been accepted as a truism, Ambridge and Lieven (2011, p. 5), for instance, vehemently reject a "third-way" or "radical middle" account. The crucial factor both positions disagree on is whether a faculty for language exists at birth or not, so there cannot be a compromise bridging the gulf between the theories. Ambridge and Lieven (2011) posit that theorists must therefore face the challenge of providing evidence in support of their standpoints and weakening respective counterclaims. Shatz (2007), too, acknowledges that the main theories of the field are polarising and often depicted as stark, incompatible opposites. Yet, this does not reflect the reality of research, which is in practice informed by previous contributions: "[P]rogress in a field typically comes via rapprochement, as both earlier questions and earlier answers are modified to accommodate more recent findings" (Shatz, 2007, p. 3). The dichotomous relationship between generativist and usage-based approaches, formalism and functionalism as well as nature and nurture is, hence, better viewed as a continuous process of mutual adaptations (Shatz, 2007, p. 3).

2.1.2. Ages and Stages: The Acquisitional Schedule

In studying language acquisition, researchers have become aware of remarkable parallels in linguistic development of children across languages and cultures (Ibbotson, 2020). Language abilities begin to develop well before birth; prenatal infants are not only sensitive to their mother's voice but also exhibit a clear preference for their first language as opposed to one with foreign acoustic properties (Kisilevsky et al., 2009). Once outside the womb, infants face the challenge of learning phonemes and segmenting the speech stream using the mechanism of categorical perception (Lust, 2006, pp. 143-144, 148). In order to do so, they are heavily reliant on caregiver input, known as infant-directed speech (IDS) or child-directed speech (CDS). Most adults adjust their style on a phonological, lexical and structural level, possibly to facilitate the acquisition of language (e.g. Kuhl et al., 1997; Cameron-Faulkner et al., 2003; Stoll et al., 2009). Children must attend to their linguistic environment to detect cues and recurring, stable sequences:

> These chunks, with more analysis, will become identifiable as units, such as words and morphemes, as well as larger expressions. Critically, therefore, they need to attach meanings to such chunks […] to make use of them – to understand them and, later on, produce them. (Clark, 2016, p. 72)

Children must not only acquire comprehension abilities but also the capacity to produce language. While children's first words around the age of 12 months are widely considered a major milestone on their language journeys, they have to be discussed in light of all the exposure and training that came before. Vocalisations early after birth are outside of infants' control and can be grouped into reflexive noises (vowel-like) and vegetative noises (consonant-like); they do not reflect any language-specific properties or communicative intentions. In the cooing stage (about 2-4 months), babies combine velar consonants [g] and [k] and back vowels [a] and [u] and experiment with noiseless lip and tongue activities (van der Hulst, 2023, pp. 349-350). With more articulatory control, children begin to engage in vocal play and produce more systematic babbles, reduplicated and variegated speech syllables (Vihman, 1996, p. 109-111). De Boysson-Bardies and Vihman (1996) compare early vocalisations in English, French, Swedish and Japanese and stress that babbling already reflects sound patterns of the target language at six months of age. Towards the end of the first year of life, researchers notice so-called melodic utterances in children's linguistic manipulations, which are characterised by rhythm and intonation but no identifiable words (van der Hulst, 2023, p. 350).

During the first twelve months, children learn to parse input into syntactic units and gain a basic understanding of the structure and order of their first language. This substantiates the claim that the perception of syntactic primitives precedes production as the first recognisable words occur at the age of one (Lust, 2007, pp. 193-194). Tardif et al. (2008) investigate the first words of children from the United States, Hong Kong as well as Beijing and observe striking commonalities across languages. Children appear to learn concrete, manipulable object nouns, action verbs or terms denoting people first. Yet, the authors also highlight some conspicuous differences which they attribute to cultural practices of naming, socialisation and identity formation (Tardif et al., 2008, p. 937). Based on data from 29 languages and 75,000 children, Frank et al.'s (2021) wordbank project demonstrates that, despite obvious individual differences, expressive vocabulary learning accelerates in the second year. Like first words, early combinations serve a holophrastic function as they express the content of whole utterances in just one or two linguistic symbols (Tomasello, 2000, pp. 65-66). A cross-linguistic study by Bowerman (1975) emphasises that syntactic development is similar in form and function during the two-word stage:

Table 1. First two-word utterances across four languages (translated to English) from Bowerman (1975, pp. 271, 274)

Two-word utterances	English	Finnish	Samoan	Luo
Agent-action	*Mommy push*	*Man dances*	*Goes Va*	*Car runs*
Action-object	*Bite finger*	*Drives car*	*Spank me*	*Eat medicine*
Possessor-possessed	*Dolly hat*	*Aunt car*	*Ball yours*	*Head mine*
Demonstrator-object demonstrated	*That candy*	*There cow*	*Ball there*	*It clock*
Adjective-noun	*Big bed*	*Little fish*	*Children older*	*Pepper hot*

Comparable results were obtained for three-term constructions (Bowerman, 1975, p. 278). Soon children's increasingly sophisticated ideas cannot be captured by merely stringing together a few words, so they start playing with more complex utterances; telegraphic speech (2;0-2;6 years) is typified by a high salience of content words, especially nouns and verbs, and the omission of function words. This illustrates that children at this stage consider grammatical elements less relevant to conveying their messages and, instead, choose to focus on items, like objects or actions, that contribute more meaning (Ibbotson, 2022, p. 111).

have briefly sketched the cognitive and linguistic skills children acquire over the first years of life. I will elaborate on construction learning and the emergence multiword utterances in the next section. Many linguists have described the stages of development; Figure 1 brings together Vihman (1996) and van der Hulst's (2023) discussion of the language learning process graphically:

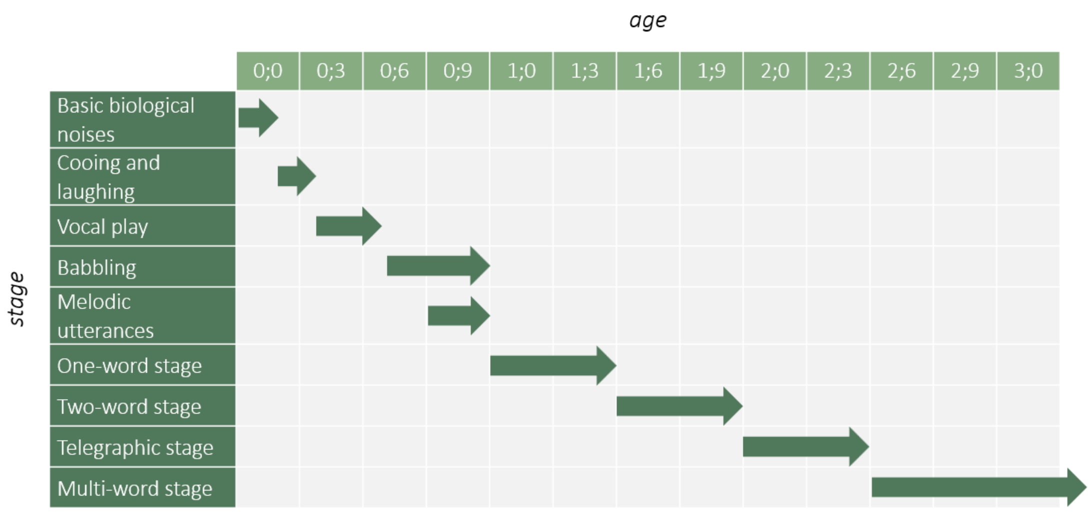


Figure 1. From sounds to words – timeline of children's early productions (summarised from Vihman, 1996; van der Hulst, 2023)

This schematic visualisation of language development invites the conclusion that the stages introduced here are strictly delimited and follow in orderly sequence. It is worth noting, however, that there is, as a matter of fact, continuity, overlap and individual differences. Some contributions have proposed alternative classifications, too; Morse and Cangelosi (2017), for example, use the universal phonetic discrimination (0;0-0;8 years), perceptual reorganisation (1;2-1;8 years), vocabulary spurt (1;8-2;0 years) and language production stage (2;0+ years) in their robotics study to the acquisitional schedule. They program the humanoid robot iCub to learn language like a child in order to simulate how children level up from one stage to the next. Contrary to common belief, which has long portrayed language learning as gradual or exponential, developmental transitions can be explained with neural and perceptual readiness which set the child up for the emergence of competences needed in the subsequent stage (Morse & Cangelosi, 2017, p. 32).

2.2. Construction Grammar: The Basics

As I already hinted at, usage-based attempts to explaining language acquisition are guided by the theoretical principles of construction grammar. Goldberg (2006, p. 14) goes even one step further in stating that "[m]ost construction grammars these days are usage-based", too. It would be a fallacy to construe construction grammar as a monolithic theory when it should, in fact, rather be treated as a set of approaches sharing the conjecture that meaning arises from language use (Ziem & Lasch, 2013, p. 36).

After serious re-evaluations of the behaviourist theory of language acquisition, which assumes that children learn language through mere imitations, Chomsky established generativism as a dominant paradigm in the fifties (Ziem & Lasch, 2013, pp. 7-8). While most linguists, at first, subscribed to an innate language faculty and the distinction between performance and competence as well as lexicon and grammar, the following decade saw more and more contributions departing from these ideas (e.g. Langacker, 1987; Lakoff, 1987; Fillmore et al., 1988). Cognitive grammar was consequently formed as a counter-project to the generativist framework at American universities and marked the onset of a new era of grammatical analysis. Construction grammar, on the other hand, belongs to the congeries of cognitive approaches and is applicable to a multitude of areas of linguistics, especially to first, second and foreign language acquisition (Koch, 2019, pp. 26-27).

Language is not learned based on abstract rules embedded in the genotype (Chomsky, 1980) but indeed following linguistic patterns with open slots extracted from input (Tomasello, 2000). Langacker's Cognitive Grammar (1987, 2008, 2013) and Goldberg's Cognitive Construction Grammar (1995, 2003, 2006, 2013) as well as their discussion of constructions as mappings of form and function have been elementary to non-nativist child language research (Koch, 2019, p. 24).

2.2.1. What's in a Construction?

Most constructivist proposals are only distantly related, yet united by "their love of interesting and complex data and […] their dislike of most work in the UG camp, whose theories they regard as distorting the basic nature of individual languages to fit a pre-conceived mold" (Sag et al., 2012, p. 2). Goldberg (2013, pp. 15-16) determines four tenets that constructivist theories share: linguistic structures are (a) defined as constructions, (b) organised in networks, (c) analysed in their surface forms and (d) examined for functional and cognitive motivations across languages. The vast majority of constructivist practitioners, moreover, describe their work as usage-based, although this label is not universally accepted (Goldberg, 2013, p. 16).

The entity at the heart of construction grammars is a construction; the term 'construction' has been used very generously in linguistics to denote different structures at large or in a narrower sense (Östman & Fried, 2004, p. 1). Since the boundaries of constructions are fuzzy and obscure, they require some clarifications. Constructions are extensions of the Saussurean notion of the ‘sign’, an arbitrary pairing of sound pattern and mental concept (or signifiant and signifié; de Saussure, 1916). Many years after de Saussure, constructivists realised that such form-meaning correspondences are useful in describing grammatical units beyond the morpheme and word as well (Hoffmann & Trousdale, 2013, p. 1). In that sense, constructions can be defined as “stored pairings of form and function, including morphemes, words, idioms, partially lexically filled and fully general linguistic patterns” (Goldberg, 2003, p. 219). Tomasello (2003, p. 99) adds that constructions stand in diametrical opposition to the rules posited in generative grammar and map onto a continuum of different levels of complexity. Crucially, they are arbitrary, symbolic and pair elements of form with semantic structure components (Croft & Cruse, 2004, p. 260):

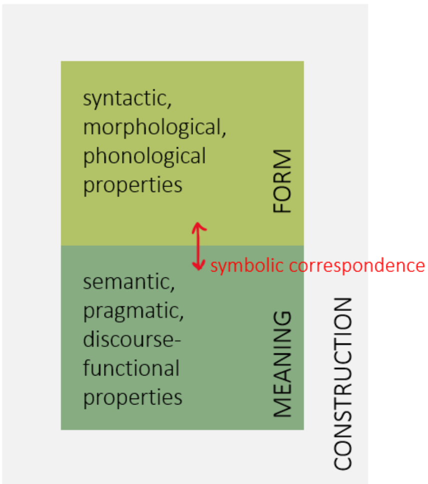


Figure 2. The symbolic structure of a construction (Croft & Cruse, 2004, p. 258)

This visualisation of a construction’s anatomy indicates that one cannot divorce form from meaning because they are intrinsically linked. In order to further explicate constructions and their role in language acquisition, it is helpful to further delineate them from other basic units of grammatical analysis and linguistic representations. These building blocks fall onto two dimensions, simple (words and categories) versus complex (constructions and expressions) and concrete (words and expressions) versus abstract (categories and constructions), as shown in the following table (Tomasello, 2003, p. 100):

Table 2. Terminology for discussing constructions at different levels of complexity and abstractness by Tomasello (2003, p. 101)

	Simple	*Complex*
Linguistic abstractions	Categories	Constructions
Concrete pieces of language	Words (morphemes)	Expressions (phrases)
Linguistic symbols	Items	Structures

Thirdly, 'items' and 'structures', which are neutral labels, pertain to the group of linguistic symbols. According to Tomasello (2003, p. 100), humans must master the full inventory of linguistic abstractions and concrete pieces of language but also structures that are a combination of both to communicate successfully (see 2.2.2.).

In 1995, Goldberg (1995, p. 4) first mentions formal and semantic non-predictability as a core criterion for constructions. She argues that patterns can only be considered constructions if they are not composable from its parts or other constructions in the grammar. Following this definition, *What did Liza buy the child?* does not qualify as a construction since its meaning is fully derivable and its structure is based off several more general constructions, such as individual words, the ditransitive construction, question construction, subject-auxiliary inversion construction, VP construction and NP construction (Goldberg, 2003, p. 221):

(1) [What did Liza buy the child?]

(2) a. *Liza, buy, the, child, what, did* constructions (i.e. words)
b. Ditransitive construction
c. Question construction
d. Subject-Auxiliary inversion construction
e. VP construction
f. NP construction

Constructions, like those listed in (2), can be joined freely to create 'constructs', that is, actual expressions (1). The distinction of construction and construct therefore boils down to one between "generalisations and concrete instances, between abstract types and the tokens that instantiate them" (Hilpert, 2014, p. 12). Goldberg's earlier diagnostics for detecting constructions have since been revised; now constructions either have to be unpredictable with regards to some aspect of their form or occur at a high enough frequency to be fully predictable and thus stored as formulae (Goldberg, 2006, p. 5).[3]

Constructional knowledge is not assembled in an unstructured manner. Each individual construction corresponds to a node in the taxonomic network of constructions and is hence organised in relation to others (Croft & Cruse, 2004, p. 262). Before that, Langacker (1987, p. 73) already characterised human grammar as an ordered inventory of conventional linguistic units that are in constant interaction with each other. To know a language is hence to know the repertoire of its constructions, varying in size and complexity, as stored in the "construct-i-con" (Hilpert, 2014, p. 13). Utterances are created by filling slots of constructions, or "utterance templates", with fitting linguistic items; children learn to do this in bits and pieces by abstracting from input strings (Ambridge & Lieven, 2011, pp. 124, 126).

2.2.2. Construction Learning

Soon, the field of developmental linguistics recognised the potential of constructions for better understanding the mechanisms of language learning. Over the last thirty years, the synergy of Cognitive and Construction Grammar and first language acquisition research was promoted, above all, by Langacker, Goldberg, Tomasello and colleagues in numerous contributions (Koch, 2019, p. 27). Tomasello (2003, p. 6), for instance, upholds that

> if adult linguistic competence is based, to a much larger degree than previously supposed, on concrete pieces of language and straightforward generalizations across them—with many constructions remaining idiosyncratic and item-based into adulthood—then it is possible that children's early language is largely item-based and yet they can still construct an adult-like set of grammatical constructions originating with these baby constructions.

Before going into detail, it is helpful to shortly return to the syntax-lexicon continuum set forth in construction grammar. This model allows constructionist grammarians to represent any grammatical structure, from words and phrasal patterns to idioms, as a construction (Croft, 2001, p. 17): Croft (2001) captures different types of constructions,

[3] Goldberg (2019, p. 7) has since presented another, more inclusive explanation of the concept of constructions which characterises them as "emergent clusters of lossy memory traces that are aligned within our high- (hyper!) dimensional conceptual space on the basis of shared form, function, and contextual dimensions". In this thesis, however, I follow her widely accepted 2006 definition suggesting that linguistic structures qualify, and are stored, as constructions if "some aspect of its form or function is not strictly predictable from its component parts or from other constructions recognized to exist […] [or] if they are fully predictable as long as they occur with sufficient frequency (Goldberg, 2006, p. 5).

three of which are particularly interesting for child language studies, using the variables of schematicity and specificity. He contrasts “complex and specific”, “complex and schematic” as well as “complex but bound” (partially schematic) constructions (Croft, 2001, p. 17), also referred to as ‘frozen phrases’, ‘lexically specific or item-based’ and ‘wholly abstract’ constructions, respectively. Item-based constructions have also been discussed under ‘schemas’ or ‘slot-and-frame patterns’ and characterised as ‘partially productive’ or ‘low-level’ (Ambridge & Lieven, 2011, pp. 126-127). Their relevance to children’s early productions has also been proven empirically (e.g. Tomasello, 1992; Dąbrowska & Lieven, 2005).

Behrens (2009, p. 433) points out that children are not immediately equipped with general language patterns but first learn to abstract across linguistic material to which they are exposed. There is a number of cognitive processes involved in advancing from rote-learned items to more flexible frames with open slots. Tomasello (2003, p. 5) names categorisation as an indispensable skill in language acquisition, encompassing abilities like perceiving similarities and forming analogies. In a later work, he adds “schematization and analogy”, “entrenchment and competition” and “functionally based distributional analysis” as core pattern-finding processes through which children construct a language (Tomasello, 2006, p. 286).

From a construction grammar point of view, the linguistic input children receive is adequately rich for them to acquire constructions; they notice regularities and make predictions based on statistical cues, which facilitates generalisations (Goldberg, 2006, pp. 70-72). When children perceive events or entities frequently enough, these events and entities leave a cognitive trace and become entrenched. This entrenchment permits children to create automatised and reliable routines, processing information faster upon re-occurrence (Langacker, 2000, p. 93). Language learners, however, must not only build habits based on their linguistic environment but also reach the right conclusions about what not to say (Goldberg, 2011, p. 132). Constructions are often competing, yet in communicative situations choices are highly motivated; preemption provides learners with the likelihood of one more appropriate form being selected over others (Goldberg, 2011, p. 151).

Schematisation and analogy, related forms of categorisation, are restricted by entrenchment and preemption (Tomasello, 2006, p. 287). Bybee (2006, p. 711) maintains that "the general cognitive capabilities of the human brain, which allow it to categorize and sort for identity, similarity, and difference, go to work on the language events a person encounters, categorizing and entering in memory these experiences". This means that children create the abstract, in the form of schemas, from the concrete, i.e. linguistic material, by filtering out differences and portraying structures at a lower granularity (Langacker, 2000, pp. 93-94). Taken together, the "abstraction of a unit is then just a matter of selective reinforcement—the progressive entrenchment of recurring configurations" (Langacker, 2009, p. 628). Langacker (2009) after all calls for a dynamic conception of linguistic structure, arguing that language units are not separate boxes but always immanent in more detailed specifications. In Figure 3, *A* represents a schema whereas *A', A'', A'''* are instantiations of said schema (Langacker, 2009, pp. 628-629):

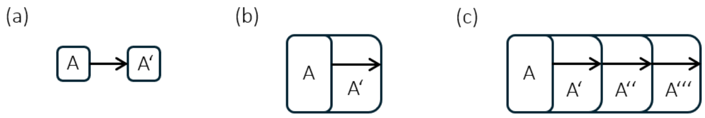


Figure 3. Immanent linguistic units (Langacker, 2009, p. 629)

Schemas are strengthened across usage events until they are readily activated (Langacker, 2009, pp. 629-630). Making such abstractions, however, is not a trivial, straight-forward task; it is not until more advanced stages of development that children begin to move away from copying input. Since they use many direct repeats, they give the impression of being productive and creative language users when their utterances are, in fact, very constrained and not as abstract as they seem (Lieven & Tomasello, 2008, p. 170).

Tomasello & Brooks (1999) study children's syntactic productivity over the first three years of life. At first, children perceive usage events holistically, do not mark participant roles, nor categorise specific scenes. Holophrases and simple word combination are lexically specific; the pivot schema stage, thus, constitutes the beginning of children's creative use of constructions (Koch, 2019, p. 136). The table below presents an updated version of Tomasello and Brooks's construction types and key properties:

Table 3. Construction types and their characteristics during early syntactic development (originally from Tomasello & Brooks, 1999, p. 164; extended by Koch, 2019, p. 99)

	Lexical partitioning of scenes	Syntactic marking of participant roles in scenes	Categorisation of specific scenes	Lexically specific
Holophrases (1;0 years)	✗	✗	✗	✓
Word combinations (1;6 years)	✓	✗	✗	✓
Pivot schemas (1;6 years)	✓	✗	✗	✗
Item-based constructions (2;0 years)	✓	✓	✗	✗
Complex constructions (3;0+ years)	✓	✓	✓	✗

From birth, children experiment with language and attempt to reproduce utterances they encounter in their linguistic environment. However, aged one, they can only produce single components of adult constructs, so-called holophrases, that are intended to capture their full meaning (Tomasello, 2000, p. 65). MacWhinney (2014, pp. 34-35) suggests that the transition from words to sentences is hardly noticeable because two-word combinations already rest on some basic understanding of syntax. Nonetheless, a convincing body of research has revealed that the learning process fundamentally changes at the age of 18 months. Like holophrases, word combinations are grounded in social scenes and convey communicative intentions. While the two elements in such constructions are usually equals in terms of status, pivot schemas, emerging at the same age, consist of one part that structures the utterance. Pivot schemas are systematic patterns with one open slot (*More X, Where's X, X it*) and can therefore be described as children's earliest abstractions (Tomasello, 2003, p. 114). Multiword productions become increasingly diverse and complex as children apply a cut-and-paste strategy to arrive at item-based constructions; they abstract across strings in a piecemeal fashion until they showcase adult-like constructions at 3;0 years (Diessel, 2013, p. 16; Koch, 2019, p. 137).

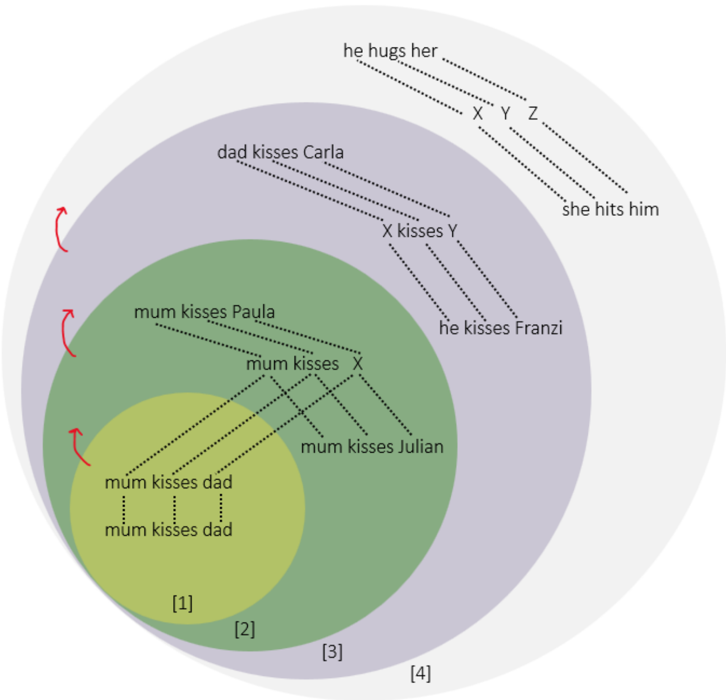


Figure 4. The schematisation process (translated from Koch, 2019, p. 139)

Koch (2019) exemplifies the schematisation process with the fictitious sentence *mum kisses dad* (originally in German), a fixed phrase that is at first memorised and produced as a whole (Figure 4, [1]). The child will, further, hear similar utterances, such as *mum kisses Paula* or *mum kisses Julian*, through which they recognise analogies and schematise the item-based positional pattern *mum kisses X* [2]. Soon, they will also notice from input strings that, like *X*, *mum* occupies a variable slot and can, thus, be replaced by other referents, for instance *dad* or *he*. *X kisses Y* [3] exhibits even less lexical specificity than earlier utterances. To form the fully abstract schema *X Y Z* [4], children have to resort to commonalities between *he hugs her, she hits him* and all earlier input in [1], [2] and [3], too (Koch, 2019, pp. 139-140).

Many linguists have taken special interest in item-based slot-and-frame patterns in language learning (e.g. Croft & Cruse, 2004; Langacker, 2008; Lieven & Tomasello, 2008; Bybee, 2010), speculating that fully abstract constructions are rather the exception than the rule: "There is reason to believe that lower-level schemas, expressing regularities of only limited scope, may on balance be more essential to language structure than high-level schemas representing the broadest generalizations" (Langacker, 2000, pp. 92-93). The role of such partially productive schemas in acquiring spatial language constructions will be assessed in the main part of this thesis.

2.3. Space Under Construction: Spatial Language

Both the study of space and the study of language have received much scholarly attention over the past centuries. At their crossroads, language *of* space, not to be confused with language *in* space, is the focus of a growing body of literature (Levinson, 1996, p. 353). Locating objects and events, interacting with them in space and talking about those situations is arguably a vital part of the human experience. Space is a tangible, semantically non-restrictive domain, which can be metaphorically extended to cover more abstract domains (Radden & Dirven, 2007, p. 303), including time (Bennett, 1975; Wierzbicka, 1993; Boroditsky, 2000) and emotion (Lakoff & Johnson, 1980). As we largely think in space-like terms, "converting non-spatial problems into spatial ones seems to be one of the fundamental tricks of human cognition" (Levinson, 2003, p. 16). It has been suggested that thinking without some degree of spatialisation and visualisation is impossible; framing problems with spatial models could thus result in enhanced processing efficiency (Johnson-Laird, 1983).

Levinson advocates for accessing mental concepts of space through language and poses the following question: "How do people refer to places, describe spatial arrangements, say where someone is going, and so forth?" (Levinson, 1996, p. 355). Spatial language research investigates "the total set of linguistic expressions that denote aspects of space" (Carlson & van der Zee, 2005, p. 1) such as "those words and simple phrases that encode objects and places" (Landau & Jackendoff, 1993, p. 218). In that sense, spatial language is used to describe location, motion or composition of various entities ranging from the close proximity of our own bodies to natural landscapes and human-made structures at a larger scale. While the field has long benefited from interdisciplinary and cross-linguistic work, due to major advances in technology, the computationally-based generation and interpretation of spatial language have recently gained momentum and promise further insights into representations of space (Stock et al., 2022, pp. 185-186).

Although structuring space through language is inherently human and a vital part of daily communication, the 7.000 languages of the world possess different ways of expressing spatial relations (Levinson & Wilkins, 2006, p. 2). Investigations have long been biased towards Eurocentric research traditions, thereby perpetuating a view of egocentric, anthropomorphic and relativistic spatial thinking. Spatial language, ostensibly, reflects this centrality of the body in space. In fact, however, substantial neurological evidence (O'Keefe & Nadel, 1978; Burgess et al., 1999) and linguistic data from non-Indo-European contexts disconfirm this claim. Instead, as numerous studies postulate (Svorou, 1994; Levinson, 2003; Boroditsky, 2012), spatial information is conveyed in different forms across languages, which is taken to confirm that there is diversity, not uniformity, in cognition, too (Levinson, 2003, p. 22). Yet, the centrality of space remains undebated; whereas only few languages have a word for the abstract concept of *space*, all have lexicalised the interrogative *where* to ask for locations and directions (Levinson, 1996, p. 359).

The English language expresses spatial relations mostly through prepositional phrases (Radden & Dirven, 2007, p. 303), but also uses adverbs, verbs, adjectives and nouns to describe, amongst others, compositions, types of movements or properties of entities in physical space (Carlson & van der Zee, 2005, p. 1). A scoping review by Wu et al. (2022), moreover, establishes that children's early spatial language is influenced by parental input and closely associated with spatial cognition development. In the next section, the cognitive linguistic foundations and relevant terminology of spatial language are laid out before discussing main insights from child language and developmental studies.

2.3.1. Language of Space in Cognitive Grammar

In the section above, I have argued for the significance of spatial concepts in language and cognition. Space serves as a template for thinking about non-spatial domains and plays an important role in cognitive linguistics as well (Talmy, 1983; Langacker, 1987). To usefully discuss spatial semantic categories, Levinson and Wilkins (2006) offer a 2-category system that distinguishes static (stasis) and dynamic (kinesis) spatial information:

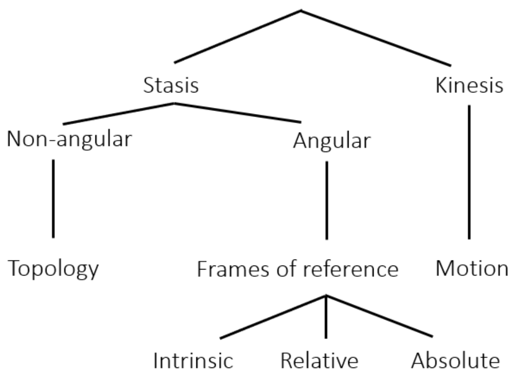
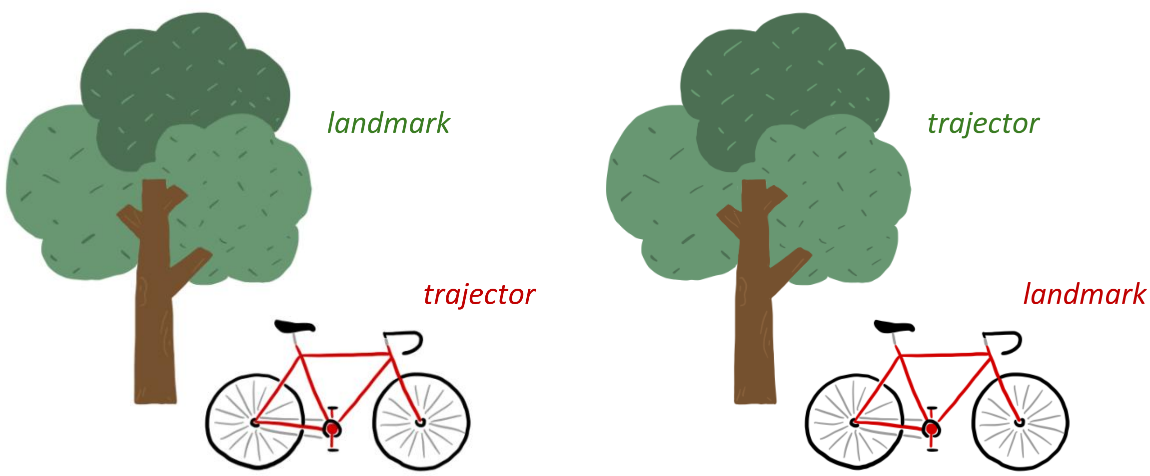


Figure 5. Conceptual subdivision of the spatial domain by Levinson and Wilkins (2006, p. 3)

These functional distinctions are dealt with in the remaining part of this section. First, the act of locating entities requires the speaker to situate them relative to one or more others which are identifiable to the hearer (Svorou, 1994, p. 8). In his paper, Talmy (1983, p. 230) maintains that

> [t]he spatial disposition of a focal object in a scene is largely characterized in terms of a single further object, also selected within the scene, whose location and sometimes also "geometric" properties are already known (or assumed known to an addressee) and so can function as a reference object. The first object's site, path, or orientation is thus indicated in terms of distance from or relation to the geometry of the second object.

He adopts the terms "figure", for the primary, focal object to be located, and "ground", to denote the secondary object that acts as a reference point, from Gestalt psychology. Figure and ground have further been discussed under "locans" and "locatum", "relans" and "relatum" or, most notably, "trajector" and "landmark", labels first coined by Langacker (1987) and now widely accepted by the research community (Svorou, 1994, p. 9). It has, moreover, been put forward that the two are unequal in status and exhibit different properties; on the one hand, trajectors (TR) are usually moveable, geometrically simpler, smaller and more salient once perceived (Talmy, 2000, p. 183) whereas landmarks (LM), on the other hand, are larger, immobile and independent entities that are backgrounded (Svorou, 1994, pp. 11-12). In *The bike is under the tree*, illustrated below (Figure 6), the *bike* is considered the TR and the *tree*, due to its fixed position and size, the LM relative to which people orient themselves.



(3) The bike (TR) is under the tree (LM). (4) ?The tree (TR) is over the bike (LM).

Figure 6. Asymmetry between trajector and landmark in a static event (my example)

These example sentences substantiate the asymmetry condition between trajector and landmark as reported by Langacker (1987, p. 231). While (3) reads naturally, (4), albeit grammatically well-formed, will strike most speakers as odd because a tree, larger and non-movable, seems a more logical reference point. Asymmetrical relations stem from differences in "size, containment, support, orientation, order, direction, distance, motion, or a combination of these [factors]" (Svorou, 1994, p. 8).

Talking about trajectors, landmarks and their relations requires a certain knowledge of the context in which these entities interact; Miller and Johnson-Laird (1976, p. 59) loosely define the "region of a thing […] as a rather indeterminate penumbra surrounding it". The notion of 'region' is vital to theories that prioritise people's relativistic experiences in space. Based on entities' structural, interactional and functional attributes, these theories differentiate between three types of regions: interior regions, exterior regions and things as regions (Svorou, 1994, pp. 15-16). Another concept relevant to the cognitive linguistic understanding of space and thus worthy of introduction is that of 'reference frames' or 'frames of reference' (RF); they are best characterised as "abstract mental structures—coordinate frameworks that organize a set of spatial relations" (Shustermann & Li, 2016, p. 116). Levinson's (2003) survey of more than 20 languages presents a tripartite typology of frames of reference, "relative", "intrinsic" and "absolute", for the world's languages, based on axes and anchoring systems.[4]

Levinson (2003, p. 38), moreover, asserts that the category of "deictic frames of reference", established in many other sources, is misleading and should be discarded in favour of his taxonomy of viewer-centred (relative) and object-centred (intrinsic) frames as well as reference systems anchored to fixed bearings originating from the ground (absolute). Instead, each reference frame can, but is not required to, have a deictic centre which links the spatial relation to the speaker's position. The examples (5) to (10) listed below demonstrate the variety of reference frames found in the languages of the world according to Levinson (2003, pp. 50-51):

[4] Since non-angular stasis is more relevant to English-speaking children's early spatial language, regions and frames of reference are only tangentially connected to the current study and thus not treated in more detail here. Svorou (1994), Levinson (2003) and Brown (2015) offer a comprehensive discussion of linguistic expressions of space, including the aforementioned concepts and their subtypes.

Table 4. The three linguistic frames of reference and their deictic centres

Examples (taken from Levinson, 2003, pp. 50-51)	reference frame	deictic centre
(5) The ball is in front of me.	intrinsic	yes
(6) The ball is north of me.	absolute	yes
(7) The ball is in front of the tree (from where I am standing).	relative	yes
(8) in front of the chair	intrinsic	no
(9) north of the chair	absolute	no
(10) in front of the tree from Bill's point of view	relative	no

Languages differ substantially in which frames of reference they foster and how readily they are available to express spatial relations. Cross-linguistic research further highlights that a certain number of them use reference frames in lieu of topological notions such as *in* or *under* (Levinson, 2003, p. 34). Boroditsky (2012), for example, discusses the striking case of spatial terms in Guugu Yimithirr; speakers of this Australian Aboriginal language prefer an absolute reference system, relying on cardinal directions to determine and describe the place objects and situations occupy at all scales. This gives rise to structures like *There's an ant on your southeast leg*, which would seem unnatural to native speakers of English (Boroditsky, 2012, p. 618). In fact, in English it is prepositions that profile the relationship between LM and TR. The location of a static entity or the direction of one in motion is specified relative to a landmark with locative and directional prepositions, respectively; another basic spatial relation is that of extent, which denotes length, width or depth (Radden & Dirven, 2007, p. 307).

Location, direction and extent can be expressed either through reference to the dimension of a landmark or the orientation between two or more landmarks. The dimension of a landmark, that is, its geometric shape, delimits the regions in which the trajectory can potentially be situated and is specified by means of dimensional prepositions. In English, these dimensional prepositions are arranged based on their dimensionality and locative or directional use (Radden & Dirven, 2007, pp. 309-310), as described in the adjacent table:

Table 5. The basic dimensional prepositions of English following Radden & Dirven (2007, p. 310)

dimensions	location		direction	
	PLACE	SOURCE	GOAL	PATH
0-dimensional POINT	*at, by, near, close to, with*	*from, away from*	*to, at, for, towards*	*by, past, via*
1- & 2-dimensional LINE / SURFACE	*on, on top of*	*off (of)*	*on(to), against*	*along, about, around*
3-dimensional CONTAINMENT	*in, within, inside, between, among*	*out of, outside of*	*in(to)*	*through, throughout*

The topology of dimension distinguishes 0-dimensional (points or indeterminacy), 1 dimensional (lines), 2-dimensional (surfaces) and 3-dimensional (containment) relations. In combination with place, source, goal and path, the dimensions culminate in a rich inventory of English dimensional prepositions (Radden & Dirven, 2007, p. 310). In contrast, one can also determine the region of an entity by means of the orientation between multiple landmarks which are mentally linked, like in *A man stood behind the curtain*. It is not imperative that these reference points are made explicit; it may well be the case that one or both remain implicit. Orientational prepositions build on the image schemas UP-DOWN (verticality), FRONT-BACK (horizontality) and LEFT-RIGHT (Radden & Dirven, 2007, pp. 313-314).

2.3.2. Language of Space in Acquisition

Over the years, much analytical interest has been taken in how children acquire a complex system of knowledge such as space while also learning a language. Two competing camps, who disagree on the respective contribution of non-linguistic spatial cognition and language towards mastering spatial categories, have emerged (Bowerman, 1996, p. 145). The cognitive approach to spatial semantic development claims that children have general concepts of space in place before spatial words emerge, thus declaring cognition the driving force (e.g. Clark, 1973; Johnston & Slobin, 1979). Bowerman (1996, p. 145), an advocator for the role of language, on the other hand, insists that the cognitive argument is incomplete and underestimates the effect of linguistic experiences: "Although nonlinguistic spatial development clearly paves the way for children to acquire spatial morphemes, learners must attend to the linguistic input to discover the particular way space is organised in their language".

Her objections rest on evidence gathered in a set of studies by Bowerman and Choi (Choi & Bowerman, 1991; Bowerman, 1996; Bowerman & Choi, 2003). In one of their experiments

(Choi & Bowerman, 1991), the researchers tested language-related differences in spatial categorisation in English and Korean 1- to 3-year-olds. These two languages diverge considerably in how they partition spatial meanings: while English makes a categorical distinction between contact situations of containment (*in*) and support (*on*), Korean subdivides containment, amongst others, into tight fit (*kkita*) and loose fit (*nehta*). These language-specific conceptual organisations are already reflected in spontaneous speech of children younger than 20 months, pointing towards a receptiveness to discriminations early on as opposed to universal concepts of space (Choi & Bowerman, 1991, p. 83). Similar results were obtained in preferential looking tasks with pre-verbal infants (Bowerman, 1996), elicitation design studies (Bowerman & Choi, 2003) as well as error analyses (Bowerman & Choi, 2003), which again suggests that learning spatial meanings does not require mapping words to preexisting categories but rather sensitivity to the categorisation system of the input language. This, therefore, presents at least the possibility that language influences non-linguistic cognitive development (Bowerman, 1996, p. 170).

Despite the evident differences in how languages encode space, the question remains if there are general developmental patterns in spatial language acquisition independent of linguistic background. Several lines of evidence have postulated that prelinguistic infants already understand spatial distinctions such as *above* and *below* (Quinn, 1994), *between* (Quinn et al., 2003) as well as contact and moving relations (Needham & Baillargeon, 1993; Roseberry et al., 2012). In preparation for their first productions, children become perceptive to more complex spatial contrasts and learn to create shared interactional spaces (Casasola & Cohen, 2002; Tomasello & Carpentar, 2007). Further investigations (Tomasello, 1987; Weist et al., 2000; Meints et al. 2002) demonstrate that simple prepositions, such as *in*, *on* and *under*, are comprehended and produced before children's second birthdays. Today, there is a consensus amongst scholars that early-acquired spatial terms are connected to topological relations of containment, accessibility, support and verticality, while words revolving around proximity and projective relationships emerge at a later stage (e.g. Piaget & Inhelder, 1956; Johnston & Slobin, 1979; Bowerman & Choi, 2003).

Grimm's (1975) study with 137 German-speaking children as well as Vorster's (1984) analysis of six longitudinal corpora in Afrikaans indicate that there seems to be a developmental priority to learn spatial prepositions before temporal and grammatical ones. Tomasello (1987) investigates the acquisition of prepositions in the second year of life and confirms that the spatial oppositions *up-down*, *on-off*, *in-out* and *over-under* emerge first. He further notices that their early occurrences were mostly non-prepositional; they were used verb-like in protophrases denoting activities, such as *up* for *pick me up*, or omitted, *draw paper* or *ball garbage*, then in increasingly complex combinations around 24 months of age (Tomasello, 1987, pp. 83-86).

Tomasello presents two potential explanations why spatial relations precede other conceptual relations. One, building on Johnston and Slobin's (1979) waiting room metaphor, the acquisition of locative prepositions could be tied to cognitive development: "[A] linguistic form enters into the 'waiting room' when the child discerns, primarily on non-linguistic grounds, a particular concept […] [and] leaves the waiting room when the child acquires a means for expressing that concept in the language s/he is learning" (Tomasello, 1987, p. 90). In fact, however, Tomasello favours a version that respects linguistic factors; early spatial propositions are used as holophrases by caregivers whereas those acquired later occur primarily in bigger phrases, in which they are more difficult to extract for the child, thus staying in the waiting room longer (Tomasello, 1987, pp. 91-92).

2.4. Interim Summary

Language allows users to produce unlimited meanings with limited building block; it is flexible and complex, so much so that children cannot possibly learn such a productive grammar from input alone (Bannard & Matthews, 2008, p. 241). Generative grammar, the dominant paradigm of the 1960s, thus proposes that humans possess an innate language faculty (e.g. Chomsky, 1965; Pinker, 1994). Yet, an alternative account is offered by usage-based linguists; they postulate that learners use general cognitive skills, like intention-reading and pattern-finding, to acquire linguistic knowledge from experience (Langacker, 1987; Tomasello, 2003; Goldberg, 2006). A bulk of research lends credence to the argument that children's early productions are less creative and more item-based than previously thought. In fact, they first imitate fixed formulae as encountered in the input, then generalise across these utterances to create more abstract schemas with open slots (Koch, 2019, pp. 136-138). Adopting key concepts of construction grammar, non-nativist theorists suggest that grammar constitutes a symbolic network of mappings of form and meaning, i.e. constructions (Diessel, 2013, p. 3). Cognitive grammar, in turn, provides a framework and the terminological toolkit to discuss the language of space. Although language systems vary in how they carve up the spatial world, space is tangible and thus central to human cognition (Talmy, 1983; Levinson, 2003; Radden & Dirven, 2007). Children are already sensitive to spatial contrast during the preverbal stage; they also acquire mental representations of space and linguistic terms which express spatial functions quite early in development (e.g. Tomasello, 1987; Quinn, 1994; Bowerman & Choi, 2003).

3. Previous Studies

Having explained the theoretical foundations of construction grammar and the usage-based approach to acquiring syntactic structures. I now provide some insights on the state-of-the-art research and then bring together theory and previous literature in the main part of this thesis.

Tomasello (1992), Pine and Lieven (1993) and Lieven et al.'s (1997) analyses empirically verify that children's early multiword speech is interconnected with previous in- and output. Their findings indicate that they learn language in a piecemeal fashion, using a slot-and-frame architecture. In recent years, the primacy of words as linguistic units has come under attack; interest has shifted to multiword units, which are more and more located at the core of linguistic representations, processing and learning (Christiansen & Arnon, 2017).

The processing advantages of multiword chunks (Wray, 2012) as well as humans' general sensitivity towards them (Arnon & Snider, 2010) are robustly documented. Theakston and Lieven (2017, p. 589), in their review of multiunit sequences in first language acquisition, portray language learning as language reuse; whether this line of reasoning holds, however, is crucially contingent on three conditions: multiword units must (a) occur frequently in language input addressed to children, (b) be observable in early speech, and (c) play a facilitating role in acquisition. This section presents selected studies arguing in favour of these premisses and emphasises the validity of multi-unit strings as building blocks of language.

Bannard and Matthew's (2008) study with 38 English-speaking children aged 2 to 3 years finds that they are faster and more accurate in repeating frequently occurring four-word sequences as compared to low-frequency ones. The authors' findings, however, were not entirely new; an item-based analysis of maternal child-directed speech by Cameron-Faulkner et al. (2003) shows that children are exposed to 52 core frames, which amount to half of the utterances they hear. 47% of those utterances begin with 1 of just 17 words, namely *what*, *that*, *it*, *you*, *are/aren't*, *I*, *do/does/did/don't*, *is*, *shall*, *a*, *can/can't*, *where*, *there*, *who*, *come*, *look* and *let's* (Cameron-

Faulkner et al., 2003, pp. 862-863). This loosely correlates with the distribution of item-based phrases across child productions, suggesting that learners predict constructions based on their frequency in their linguistic environment: “[T]he acquisition of a language is the acquisition of a skill, and skills are crucially dependent on repetition and repetition with variation” (Cameron-Faulkner et al., 2003, p. 868).

Lieven (2010, p. 2554), too, argues for the importance of input frequency for establishing schemas and maintains that “[b]ecause the child’s system changes and becomes more abstract, frequency effects should be at more phonologically and lexically-specific levels to start with, changing to broader scope categories with developmental time”. Empirical support for this statement comes from Peters (1977), for instance, who studies one boy’s first productions from age 0;7 to 2;3, noting that he applies a “Gestalt strategy” to language learning. The researcher identifies multiword chunks modelling adult-like intonation patterns, like [ˈóbɛ dʌ dɛ́] (*open the door*) and [síli, ìnɪ?] (*silly, isn’t it?*), well before the child uses *open*, *door* or *silly* in other contexts, let alone masters full sentences (Peters, 1977, pp. 564-566). In her later works, she concludes that children initially utter unanalysed multiunit strings, which contain a number of individual words, before segmenting them into their parts; this process is often overlooked by linguists in search of identifiable words (Peters, 1983).

Dąbrowska (2000) further illustrates children’s early preference for lexically-specific templates en route to more abstract linguistic knowledge by examining 1439 questions produced by a girl across her first years of life. Her account suggest that children use rote-learned formulae at earlier stages of development, gradually extracting schemas, like *What’s THING doing?* and *What’s THING PROCESS-ing?*, until they attain full productivity (Dąbrowska, 2000, pp. 98-99). Goldberg et al.’s (2004) corpus study also highlights that the early language learning process is item-based and rather conservative in nature, more so than previously expected. The authors explore 28-month-olds’ use of three construction types, VL (= (Subj) V $\text{Obl}_{\text{path/loc}}$), VOL (= (Subj) V Obj $\text{Obl}_{\text{path/loc}}$) and VOO (= (Subj) V Obj Obj2), to characterise how children learn argument structure generalisations. More specifically, their analysis focuses on verbs and yields insightful results concerning form-meaning associations: a single verb accounts for the majority of exemplars for each construction type. The verb *go* is crucially linked to VL constructions, whereas VOL occurs predominantly with *put* and ditransitive VOO with *give*. They explain that children appear to mirror general-purpose verbs frequent in their linguistic environment, which are foundational for acquiring constructions (Goldberg et al., 2004, pp. 295-297). According to Goldberg et al. (2004, p. 305), young learners build syntactic and semantic generalisations on the basis of abstracting over instances and continually integrating novel ones into the patterns.

Indeed, a growing body of research confirms that language input consists of many multi-unit strings and frequent frames, assisting children in grasping grammatical patterns (e.g. Lieven et al., 1997; Cameron-Faulkner et al., 2003; Goldberg et al., 2004). Stoll et al. (2009) compare Russian, German and English data to conclude that continuous frames are equally relevant in all three languages, despite obvious differences in flexibility of word order and richness of the morphological system. The considerable degree of repetitiveness allows learners to acquire the basic constructional types but also grammatical categories. Mintz (2003) evaluates the relevance of non-continuous frames for word classification in two experiments; he reasons that frame-like information constitutes solid ground on which children create grammatical categories. Slots in frames like *you__it, the__is* or *are__going* were accurately filled with material belonging to the correct class in close to all cases (Mintz, 2003, p. 102, 115). Mintz, thus, proposes distributional bootstrapping, not semantic bootstrapping (e.g. Pinker, 1984; see Mintz, 2003, pp. 110-113 for discussion), as a viable mechanism for early word categorisation.

It is evident that language-acquiring children are heavily influenced by linguistic exposure. In fact, their own productions reflect features of speech directed towards them. Behrens (2006) considers the input-output relationship in one German corpus, following a family and their child from the age of 1;11 to 4;11. She underscores that the relative parts-of-speech frequency is stable in caregiver speech across time. While the child's early structures exhibit a clear dominance of nouns and adverbs, the adult system of production rates is approximated steadily; a comparable distribution is achieved around 2;7 years (Behrens, 2006, p. 16). In light of the large quantities of input, Behrens (2006, p. 22) holds that child language actually develops slowly, at first, and is not particularly productive, a claim that is also backed by Lieven et al. (2003). The latter look at 295 multiword utterances by a 2;1-year-old and reveal that 186 of them are exact repeats either of what the child heard or produced in the previous six weeks. The rest (37%) are new constructions, although more than half of them were formed through only minor changes (Lieven et. al, 2003, pp. 343-344).

Lieven et al.'s (2003) study is also essential to this thesis because it marks the first application of the 'traceback method'. The traceback method, "a mechanical procedure of matching each novel utterance to prior utterances on the basis of number and order of shared morphemes and the frequency of exemplars" (Lieven et al., 2003, p. 359), determines the degree of relatedness between utterances. It therefore builds on the usage-based assumption that learners' linguistic knowledge becomes increasingly complex and abstract through modifications to earlier productions.[5] The researchers were able to reconstruct novel structures by substituting, adding, dropping, inserting or rearranging lexical material, although these strategies occurred at varying frequencies and were combined at times (Lieven et al., 2003, p. 362).

With slight methodological adaptations, Dąbrowska and Lieven (2005) adopt the same approach to studying syntactic questions. They aim at tapping into learner's grammatical knowledge of question constructions and their internal constituency and hypothesise that children make use of a lexically specific interrogative schema with open slots. Data from two dense corpora highlights that 9 in 10 questions were accounted for by tracing back to earlier productions. The paper also takes a developmental perspective and re-assesses the novelty or creativity of the children's production at age 3;0. While the children, at 24 months, construct about two thirds with just one operation, the number of constructions requiring two or more rises considerably with added linguistic experience in the following year (Dąbrowska & Lieven, 2005, p. 451). The fact that most questions arose from copied or re-assembled linguistic elements is cognitively plausible and confirms that children, after predominantly using formulaic strings, abstract across usage events to become less bound as far as slots are concerned (Dąbrowska & Lieven, 2005, pp. 456, 462).

A formative traceback study of four children by Lieven et al. (2009) largely reproduces findings from earlier explorations; it finds antecedents for 58-92% of utterances at 2 years of age, necessitating either no or just one operation to match earlier structures. The linguists also define six semantic slots that have been identified as recurrent in early multiword speech: REFERENT (REF), PROCESS (PRO), ATTRIBUTE (ATT), LOCATION (LOC), DIRECTION (DIR) and UTTERANCE (UTT) (Lieven et al., 2009, p. 486). Referents represent the most frequent slot fillers in the data discussed; at first, children preponderantly use simple nouns or pronouns, which are combined with determiners, possessives and adjectives with increasing mean length of

[5] In this section, I zoom in on the findings generated with the traceback approach. See 4.2. for a more detailed description of the methodological and analytical steps taken in the respective studies.

utterance (MLU). This suggests that young learners of English can assimilate noun phrases into schematised patterns well before verbs and other categories (Lieven et al., 2009, p. 504).

Both the presence of discrepancies in earlier traceback research and the rich opportunities opened up by computational approaches inspire Kol et al. (2014) to implement an algorithm of the traceback model. Their computer programme successfully relates every sixth to ninth utterance to constructions the child has produced or heard before (Kol et al., 2014, p. 187), a range comparable to Lieven et al. (2003), Dąbrowska and Lieven (2005) and Lieven et al. (2009). Traceback studies have long been biased towards monolingual English data; more recent ones, however, have examined Italian (Miorelli, 2017) and German child language (Koch, 2019) as well as bilingual codeswitches (Gaskins et al., 2019; Quick & Hartmann, 2021), too, which obtained similar results. All of them seem to agree with the claim that children's early multiword speech is recycled from earlier productions; according to Lieven et al. (2003, pp. 334-335) "rote-learning forms one end of the continuum and the generation of utterances from abstract categories and rules forms the other, with a variety of semi-formulaic patterns lying somewhere in between".

Whereas much work in linguistics focuses on how children move from syllables to words and full linguistic competence, Arnon (2021, p. 940) postulates an additional coarse-to-fine development from larger un- or underanalysed strings to smaller, segmented ones. Based on evidence, such as that discussed above, she formulates the "Starting Big approach" to language acquisition:

> Children's ability to produce or comprehend words and morphemes will be modulated by the larger linguistic context they appear in. Linguistic knowledge is not seen as an all-or-nothing state, but as a gradient one where the ability to use the correct form depends (among other things) on the immediate linguistic context. That is, children's ability to use certain morphemes and words will reflect and be impacted by multiword information. (Arnon, 2021, p. 945)

Words and multiunit structures, consequently, interact with each other and are processed in complex networks of linguistic representations. This emergentist, single-system view on language learning affects how the study of language explains mechanisms of form-meaning acquisition (Arnon, 2021, p. 945). In this section I hope to have provided ample evidence in support of Theakston and Lieven's (2017) three conditions mentioned at the beginning of this review. The importance of larger strings, slots and constructional schemas for young learners seems to be well established in a usage-based framework. Yet, there is a paucity of literature at the intersection of multiword constructions and the acquisition of language of space.

Several studies explore children and parents' natural spatial language productions in everyday situations. Ferrara et al. (2011), for instance, demonstrate that guided play with construction materials encourages both the preschoolers and their caregivers to use a wider variety of spatial terms at a higher frequency; they assign each term one of six from the University of Chicago spatial language coding system (Cannon et al., 2007 cited in Ferrara et. al, 2011). Other contributions, including Levine et al. (2012) and An and Wu (2019 cited in Yang & Pan, 2021), opt for a more simplistic taxonomy, whereas Yang and Pan's (2021) study on Chinese spatial language during block play applies Ferrara et al.'s (2011) classification. The respective systems of categories are juxtaposed in Table 6:

Table 6. Spatial language classifications according to different authors

Ferrara et al. (2011)	Levine et al. (2012)	An & Wu (2019 cited in Yang & Pan, 2021)	Yang & Pan (2021)
(1) spatial locations, (2) deictic terms, (3) dimension, (4) spatial features or properties, (5) shapes, (6) spatial orientations or transformations	(1) dimensions, features, and shapes of objects, (2) orientation and transformations, (3) location and direction	(1) spatial locations, (2) spatial tendency words	(1) spatial locations, (2) deictic terms, (3) dimension, (4) spatial features or properties, (5) shapes, (6) spatial orientations or transformations

In analysing more than 200 kindergartners' spatial talk, Yang and Pan (2021) discover a clear tendency towards the use of spatial locations and deictic terms for their subjects. They further emphasise that cooperative and creative play enhance children's exposure to spatial language and, in turn, promotes spatial thinking (Yang & Pan, 2021, pp. 13-14). These studies have strong educational implications; as spatial terms help children in organising the space around them, their development has repeatedly been linked to mathematical and spatial skills later in life (Miller et al., 2016).

Although the significance of spatial terms in child development is undisputed and Lieven et al. (2009) even name 'location' and 'direction' as productive slots in early multiword speech, further research into the cluster of constructions involving the linguistic representations of space is needed. Studies to date have, furthermore, not applied the traceback method to the study of children's spatial language, a gap this thesis intends to fill.

4. Methodology

Language acquisition research often falls back on conversation data by children of different ages, primarily in interaction with their caretakers in their homes, as compiled through corpus-based approaches. Today, corpora are fundamental resources to studying early linguistic development (Diessel, 2009, p. 1197).

According to Ingram (1989, p. 7), the history of child language studies can be divided into the periods of diary studies, large sample studies and longitudinal studies. With heightened interest in early childhood, diary studies emerged around the end of the 19th century and reported not only on linguistic but also physical, cognitive and social development. Researchers predominantly observed their own children over long stretches of time to gather insights on their general growth without apparent foci (Behrens, 2008, pp. xii-xiii). Famous examples include Darwin (1877) for English, Baudouin de Courtenay (unpublished; discussed in Smoczynska, 2001) for Polish and Stern and Stern (1907) for German child language, to name just a few.[6] Some decades later, diary studies were superseded first by bigger, cross-sectional projects and then by more meticulous longitudinal ones. As opposed to the earliest longitudinal diary studies, more recent investigations did not build on parental analyses, but data collected with the help of technologies which are now considered indispensable to the study of language acquisition. The new recording medium further allowed scholars to examine more complex structures that diaries could not capture successfully (Diessel, 2009, p. 1199).

Slobin (2014), too, argues that many linguistic discoveries were only enabled by the advancement of the trade's tools. While earlier investigations were heavily reliant on diary keeping and researcher intuitions, the introduction of innovations such as the audio and video-recorder as well as the computer brought about the availability of extensive quantities of data and more systematicity in approaching child development (Slobin, 2014, pp. 2-3). However, along with these obvious affordances, came methodological challenges with regards to

[6] Diary studies did not disappear entirely but experienced a revival in form of topical diaries with narrower research interests in specific aspects of linguistic development, such as causation (Bowerman, 1974) or verbs (Tomasello, 1992), for example. More advanced recording tools resulted in more precise and focused investigations (Behrens, 2008, pp. xiii).

annotation and transcription; Ochs (1979) criticises the lack of consensus on conventions in developmental linguistics. She recommends consistent transcription practices permitting researchers to draw the right conclusions about language-acquiring children's naturalistic speech (Ochs, 1979, p. 44).

As longitudinal studies expanded rapidly and calls for standardisation became louder, MacWhinney and Snow (1985) built an extensive computerised archive of child language transcripts called CHILDES (CHIld Language Data Exchange System). CHILDES includes its own analysis programme (CLAN) and also sets the standard for data collection and transcription practices (CHAT), incorporating many of the demands voiced by Ochs (1979), for instance. In fostering data sharing in child language research, the corpus makes the discipline a more efficient and level playing field (Diessel, 2009, p. 1199). In this thesis, I, too, will work with naturalistic language data already available online in CHILDES. My goal is to learn more about how children come to be productive and creative language users and comprehend the mechanisms of acquisition more thoroughly.

By focusing on schemas of spatial organisation, a set of structures that has, to date, received scant attention in traceback studies, I seek to contribute to the ongoing discussion of spatial terms in early childhood. This project is, therefore, informed by and intends to answer the following three clusters of research questions. First, like several researchers before me, I aim at making visible the relatedness of child language and their linguistic environment:

RQ1: How many of children's spatial language utterances can be accounted for when tracing 2;6-year-olds' multiword strings back to previous productions or caretaker input? How many of them are novel and can, therefore, be considered creative language use? Which operations are applied to arrive at those novel constructions?

Hypothesis I: Previous traceback studies have found that 90% (Dąbrowska & Lieven, 2005), 85-90% (Lieven et al., 2009), 60-89% (Kol et al., 2014), 82% (Miorelli, 2017) and 85-95% (Koch, 2019) of children's earliest multiunit structures are closely related to what they heard and produced previously (note: variations are linked to age, individual differences and methodological decisions). I expect my analysis to yield results within a similar range.

Hypothesis II: A substantial number of child productions will be identical repeats of earlier input or their own output and must be treated as rote-learned frozen formulae. Most of the target utterances, however, will be constructed with low-level schemas that are only partially productive and not yet fully abstract generalisations (e.g. Langacker, 2000; Dąbrowska & Lieven, 2005; Lieven et al., 2009; Koch, 2019).

Hypothesis III: Following Koch (2019), I assume that the learners will use the combinatorial strategies of ADD, SUBSTITUTE and SUPERIMPOSE. Both Lieven et al. (2009, p. 494) and Koch (2019, pp. 204-205) show that only 5% or 6% of utterances, respectively, are derived by adding linguistic material, suggesting that language use is predominantly lexically specific at the age under investigation.

Hypothesis IV: The higher the MLU, the more multi-operation utterances and less exact matches the child produces (Lieven et. al, 2009; Koch, 2019).

Secondly, I will examine the schemas underlying children's spatial language to gain a better understanding of how linguistic material is reused or recombined to productively form novel constructions. I, therefore, ask

RQ2: Which slot-and-frame patterns are relevant to the study of the language of space in the third year of life? In which larger constructions are they used? What can be said about their internal constituency?

Hypothesis I: 'Location' and 'direction', identified as productive slots by Lieven et al. (2009), will presumably feature prominently in my dataset as well.

Hypothesis II: Other research (e.g. Ferrara et al., 2011; Levine et al., 2012; Yang & Pan, 2021) has proposed a variety of spatial terms beyond spatial locations and directions, such as orientations, transformations, dimensions, features and shapes. I therefore hypothesise that REFERENT, PROCESS and ATTRIBUTE slots will also play a role in conveying spatial information.

Hypothesis III: Since the children discussed in my study are not yet fully competent speakers, I hypothesise that the spatial schemas are used rather conservatively and stereotypically with a high degree of repetition (Goldberg et al., 2004; Dąbrowska & Lieven 2005).

Koch (2019, p. 156) urges linguists to consider the results of traceback studies with their core assumptions in mind, making the method itself object of his investigation. On a methodological level, I am thus interested in this third research question in order to evaluate the usefulness of addressing spatial language acquisition through this lens:

RQ3: What are the limitations and potentials of taking a traceback approach to studying early spatial terms and schemas?

Hypothesis I: The literature finds that not all strings can be accounted for and distinguishes lexical and syntactic fails, in which either relevant words or component parts cannot be detected in preceding interactions (Lieven et al., 2009). By extension, not all spatial language utterances will be derived successfully from antecedents.

Hypothesis II: Fails to derive certain utterances might occur due to methodological constraints inherent in the traceback approach (Koch, 2019).

In sum, this thesis explores schemas pertaining to the linguistic representation of space in early multiword speech, which will potentially be valuable for caregivers, educators and policy maker aiming to support children in their language learning and, ultimately, literacy and spatial skills. By analysing children's inventory of spatial constructions, I hope to contribute to the larger debate of usage-based first language acquisition.

4.1. Corpus and Data Selection

In order to address the aforementioned research questions, I capitalise on data provided to CHILDES, the largest open-source storage venue for child language, by collectors. It was developed in the 1980s with the intention to proliferate systematic provision of data, thus laying a solid foundation for acquisition research to thrive (MacWhinney & Snow, 1985, pp. 272, 275). So MacWhinney, Snow and their research teams at Carnegie-Mellon University in Pittsburgh and the Max-Planck Institute for Psycholinguistics Nijmegen (1985, p. 272) created a computerised database, which satisfies "(1) the need for greater efficiency in data sharing, (2) the need for increased precision in data collection, coding, and analysis, and (3) the need for increased automation of coding and analysis".

CHILDES is structured into three component parts that can be used individually or in combination with each other. First, the database itself is home to a vast number of corpora contributed by researchers to test new hypotheses (MacWhinney & Snow, 1990, pp. 459-460, 467; Corrigan, 2011, p. 274). Today, said database includes monolingual speech samples from a host of languages as well as transcripts of children from bilingual or clinical populations (Behrens, 2008, pp. xviii-xix). Most of them have been formatted using the CHAT (Codes for Human Analysis of Transcripts) notation system that is argued to account for the rich nuances relevant the study of early speech (MacWhinney & Snow, 1990, pp. 465-466). The third tool is

CHILDES's own package of software programmes, CLAN (Child Language ANalysis), with a set of search queries and options to tailor the analysis to specific interlocutors, linguistic features and utterances (MacWhinney & Snow, 1990, p. 467). Due to its many affordances, CHILDES has grown to be not only the central collection system of child language texts but also a valuable workbench (Diessel, 2009, p. 1199).

The present study analyses the early multiword strings of three English-speaking children, Eleanor, Fraser and Thomas, at 2 and a half years of age. The transcripts stem from two separate corpora, the MPI-EVA-Manchester Corpus (Lieven & Goh, n.d.-a; https://childes.talkbank.org/access/Eng-UK/MPI-EVA-Manchester.html) and the Thomas Corpus (Lieven & Goh, n.d.-b; https://childes.talkbank.org/access/Eng-UK/Thomas.html), both of which have been donated to CHILDES by principal investigators Elena Lieven and Jeannine Goh. They are publicly accessible on the platform under the subdirectory Child Language Banks/CHILDES/English-UK/.

Despite originating from different research projects, the overall conditions for data collection are stable for all children and thus allow for comparison. The families are situated in Manchester, UK, and come from similar socio-economic backgrounds. The speech samples are, furthermore, not elicited but gathered in naturalistic settings, the children's homes, during playtime with their caregivers. Lieven and Goh also used the standard Sony mini-disc recorder and Sennheiser evolution radio microphones throughout to capture the sessions in audio, permitting child and parent to move around freely (Lieven & Goh, n.d.-a; Lieven & Goh, n.d.-b). The traceback method defines a test corpus, i.e. a set of child utterances to be derived from previous discourse, and a main corpus, i.e. all productions of the recording period taken into account in the procedure (e.g. Hartmann et al., 2021, p. 230; see section 4.2.). For the test corpora, I decided upon one recording at the same age of 2;6 for each child, Eleanor 020600a, Fraser 020601a, and Thomas 020602;[7] following the argument I developed in sections 2 and 3 of this thesis, I assume the children to become increasingly creative while still adhering to partially schematic frames at this stage of language production.

In selecting recordings for the main corpora, I essentially designed my own small corpora on the basis of the existing ones. The main corpora consequently contain between 16 (Eleanor,

[7] I chose the recording Thomas 020602 as test corpus and omitted 020600 and 020601 because the family was on holiday, which changed the otherwise consistent recording parameters and impacted audio quality.

Thomas) and 20 sessions (Fraser), spanning approximately eight weeks before the test corpora were recorded. In other words, they comprise between 8 and 10 hours at the beginning of the children's 28th and 29th month of life, which amounts to 16:05:29 hours for Eleanor, 20:54:12 hours for Fraser and 18:18:04 hours for Thomas. Key information about the datasets I am working with is summarised in the table below:

Table 7. Characteristics of the linguistic material studied in this thesis

	Eleanor	Fraser	Thomas
CHILDES Talkbank	MPI-EVA-Manchester Corpus by Elena Lieven and Jeannine Goh	MPI-EVA-Manchester Corpus by Elena Lieven and Jeannine Goh	Thomas Corpus by Elena Lieven and Jeannine Goh
Test Corpus	020600a.cha	020601a.cha	020602.cha
Setting	free play at home, Manchester	free play at home, Manchester	free play at home, Manchester
Participants	child, mother	child, mother	child, mother, father
Length	1:04:35	1:02:15	1:00:47
Main Corpus	020400a, 020400b, 020401a, 020402a, 020402b, 020404a, 020404b, 020405a, 020500b, 020501a, 020501b, 020503a, 020503b, 020504b, 020505a, 020506a	020401, 020402a, 020402b,020403, 020404a, 020404b, 020405a, 020405b, 020406a, 020406b, 020500a, 020500b, 020501a, 020501b, 020503a, 020503b, 020504a, 020504b, 020506a, 020506b	020400, 020401, 020402, 020403, 020406, 020407, 020408, 020409, 020410, 020500, 020503, 020504, 020505, 020506, 020507, 020510
Setting	free play at home, Manchester	free play at home, Manchester	free play at home, Manchester
Participants	child, mother, father, grandmother, grandfather, visitor, investigator	child, mother, father, brother, visitor, investigator	child, mother, father, visitor, investigator
Length	16:05:29 (16 recordings)	20:54:12 (20 recordings)	18:18:04 (16 recordings)

The respective main corpora are by definition not dense databases since they do not cover the 5 to 20 percent of children's active and passive linguistic experience Lieven and Behrens (2012) advocate for. Tomasello and Stahl (2003, p. 119) however remark that sampling regimes should be chosen with the phenomenon studied in mind; rare ones require denser samples than those occurring at higher frequencies. Since spatial terms are ostensibly not uncommon in child speech and dense datasets are associated with a heavy workload, this traceback study will work with data collected through interval sampling, an alternative proposed by some researchers (Tomasello & Stahl, 2003; Rowland et al., 2008).

Both Eleanor and Fraser were recorded for 2 hours per day for the first week of each month; this dense period was followed by a larger temporal gap before the children were again recorded for the duration of a week at the beginning of the subsequent month. This allowed the investigators to comprehensibly capture their language use and development at less cost (Lieven & Goh, n.d.-a). In contrast, the Thomas Corpus initially consists of daily speech samples (Lieven & Goh, n.d.-b); in order to approximate the volume of the other two children, I only included the first speech samples at the age of 2;4 and 2;5 years. To allow for further comparison between the children, I measured their mean length of utterance, the ratio of morphemes (not words) over utterances, at age 2;6 using the CLAN analysis programme and the command `mlu +t*CHI file.cha`. Considering her age, Eleanor is a very advanced speaker with an MLU of 6.243; Fraser's MLU value was determined at 5.480, whereas Thomas exhibits the lowest MLU of 3.400.[8]

4.2. The Traceback Procedure

Numerous traceback studies have been conducted in recent years after the method was established in the early 2000s, most notably by the research teams of Lieven and Tomasello. The approach sets out to empirically substantiate the usage-based assumption that language learning and use are characterised by lexical specificity and was hence "originally designed to demonstrate that virtually all utterances in early child language can be attributed to a relatively small set of units that are frequently attested" (Hartmann et al., 2021, p. 230). To this end, researchers split their dataset into two parts: the 'test corpus', usually the last one or two recordings, and the 'main corpus', all previous sessions of the recording period (Lieven et al., 2009, pp. 485-486). The child productions of the test corpus, also referred to as 'target utterances', are then traced back to related ones in the main corpus. Strings are flagged as

[8] There are arguments in favour of calculating MLU in words over utterances (MLUw) instead of morphemes over utterances (MLUm) as is done here (for discussion see Brown, 1973; Parker & Brorson, 2005). Since I am only interested in the children's level of language development relative to each other, I refer to MacWhinney (2025), who proposes MLUm as the default measure in his CLAN manual.

related if they share lexical material, that is, 'component units' from which they can be derived (Dąbrowska & Lieven, 2005, pp. 446-447).

The traceback approach investigates two types of component units;[9] some structures, 'fixed strings' or 'chunks', are verbatim matches for previous structures, whereas 'slot-and-frame patterns' or 'lexically specific schemas', on the other hand, are more variable in some positions. More abstract constructions are not dealt with in traceback since children are believed to acquire language starting with concrete linguistic items and only local generalisations (Hartmann et al., 2021, p. 231). Dąbrowska and Lieven (2005, p. 447-448) suggest that slots in larger structures are formed when different elements are used to fill a frame's open position. Based on distributional grounds, Lieven et al. (2003) stipulate that any slot could be fitted with any string; later works, however, have rejected this as linguistically implausible and specified that only placing material of the appropriate semantic category would lead to slot-building. Hartmann et al.'s (2021) meta-analysis of traceback studies gives a succinct overview of their procedure and relevant features. The visualisation below depicts the process of tracing back schematically and mentions parameters along which studies to date differ:

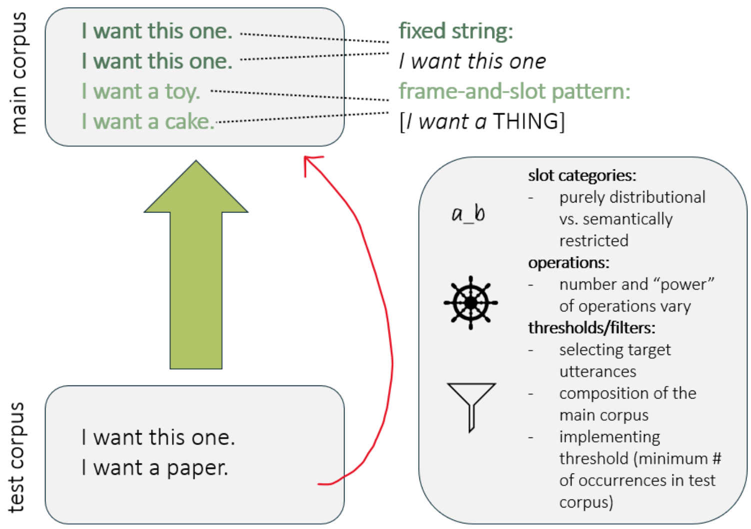


Figure 7. Illustration of the traceback approach (Hartmann et al., 2021, p. 232)

While, since its establishment, traceback has been recognised as a successful approach for investigating children's construction learning, it has at the same time been criticised for a lack

[9] Miorelli (2017), an exception to the rule, introduces two more component units to account for the flexible word order in Italian: 'fully-specific-packets' and 'schematic-packets'. For example, the strings *il pane voglio* and *voglio il pane*, different linear orders, contribute to the schematic-packet *Voglio + THING_WANTED* (Miorelli, 2017, Volume II, p. 50).

of clear demarcation (e.g. Kol et al., 2014). Existing papers have, as Figure 7 as well as Table 8 suggest, adapted the traceback method in various regards to enhance its explanatory power:

Table 8. Comparing selected traceback studies and their features

Study	Dataset	Traces back to	Operations discussed
Lieven et al. (2003)	1 English speaking girl aged 2;0, 5h/week for 6 weeks	CS and CDS	SUBSTITUTE, ADD ON, DROP, INSERT, REARRANGE
Dąbrowska and Lieven (2005)	2 English speaking children, 5h/week for 6 weeks at ages 2;0 and 3;0	CS and CDS	JUXTAPOSITION, SUPERIMPOSITION
Lieven et al. (2009)	4 English speaking children aged 2;0, 5h/week for 6 weeks, one of them full year	Only CS	ADD, SUBSTITUTE
Kol et al. (2014)	5 English speaking children, various ages 1;6 to 4;0	3 computational models: (a) CS and CDS, (b) only CDS, (c) only CS	JUXTAPOSITION, SUPERIMPOSITION
Miorelli (2017)	1 Italian speaking boy aged 2;1, 5h/week for 6 weeks	CS and CDS	JUXTAPOSITION, SUPERIMPOSITION
Koch (2019)	4 German speaking children aged between 2;0 and 2;6, 5h/week for 7 weeks	CS and CDS	SUPERIMPOSE, SUBSTITUTE, ADD
Quick & Hartmann (2021)	2 German-English bilingual children, 2;4 to 3;10 and 2;3 to 3;11, 2.5h/week	CS and CDS	SUBSTITUTE

As this table shows, these studies vary particularly with regards to dataset, constructions they trace back to, child speech (CS) or child-directed speech (CDS), as well as possible operations. Evidently, the number of children studied varies from one to five and the period from a few weeks to several months. Even with such a limited number of subjects and restricted timeframe, it is admittedly close to impossible for researchers to account for all of children's linguistic experiences; yet the research community seems to agree that the sampling intervals for naturalistic studies in L1 acquisition should be dense, which amounts to roughly 10% of children's utterances or five hours of recording per week (Rowland et al., 2008).

Another crucial decision researchers have to make concerns which strings they regard as realistic precedents. In a similar vein, Hartmann et al. (2021, p. 237) cover the selection criteria for main corpus and test corpus under 'thresholds/filters'. Except for Dąbrowska and Lieven (2005) and Quick and Hartmann (2021), who focus on question constructions or bilingual utterances, traceback investigations usually consider all child productions in the test corpus. Whether researchers derive utterances from child language alone or also include caregiver constructions in their main corpus, however, inevitable impacts the outcomes, too. There are strong arguments in favour of both approaches. Lieven et al. (2009, p. 484), among others, compose a main corpus exclusively of child productions, declaring that "we then know that the string is (or rather was, when it was uttered) part of the child's linguistic representation and this obviously makes the comparison between present and prior utterances much more straightforward". Their results should be read as confirmation that child utterances themselves are highly interconnected. Koch (2019), however, makes a point in taking both the children's and adults' multiword strings into account; he argues that language input is a fundamental resource through which learners acquire a language and must therefore relate target utterances thereto as well (Koch, 2019, p. 173; see also section 3).

Thirdly, traceback studies differ in the number and type of operations used to arrive at derivations. The earliest one, Lieven et al. (2003), identifies five operations that children apply to form novel utterances: SUBSTITUTE, ADD ON, DROP, INSERT and REARRANGE. They successfully reconstruct all strings produced by their subject, a finding that was later challenged by fellow linguists (e.g. Dąbrowska & Lieven, 2005, p. 439; Lieven et al., 2009, pp. 483-484). Lieven et al. (2003) do not define a limit to those operations so, in principle, one could derive any utterance by combining the five operations (endlessly) as desired. In the same manner, the authors relate the production *What's that funny drawing down there?* to *What's that lying down?* through three operations (Lieven et al., 2003, p. 351), which appears to be rather far-fetched:

(11) Attested utterance in main corpus: *What's that lying down?*
→ SUBSTITUTE *funny* for *lying*
→ INSERT *drawing*
→ADD-ON *there*
Target utterance in test corpus: *What's that funny drawing down there?*

This, however, is arguably not the purpose of the traceback approach; subsequent studies consequently focus on just JUXTAPOSITION and SUPERIMPOSITION (Dąbrowska & Lieven, 2005; Kol et al., 2014; Miorelli, 2017) or, under different labels, ADD and SUBSTITUTE (Lieven et al., 2009). SUPERIMPOSE and SUBSTITUTE are operationalisations of the idea that slot-and-frame pattern shape early language use, whereas ADD and JUXTAPOSE describe the linear stringing together of lexical material (Hartmann et al., 2021, pp. 233-234). Koch (2019) conclusively argues that there is a pivotal difference between SUPERIMPOSE and SUBSTITUTE; the former refers to operations where the schema and the component unit that serves as a slot filler overlap, like *shall I PROCESS* + *I open that* > *shall I open that* (Dąbrowska & Lieven 2005: 443), while SUBSTITUTE covers only instances in which the component part fits into the fitting open slot, for example, *das ist ATTRIBUTE* + *hoch* > *das ist hoch* (Koch 2019: 180).

It is important to remark that not all researchers insist that SUBSTITUTE, SUPERIMPOSE and ADD are the exact cognitive mechanisms by which children create novel utterances. In fact, Lieven et al. (2009, p. 502) "are not suggesting that the children's utterances are actually constructed by the operations given here, nor that the schemas and fixed strings that are identified are necessarily present in the child's linguistic representations". Hartmann et al. (2021, p. 233) recommend viewing traceback as a method to "reverse-engineer […] cognitive processes [and] reconstruct the 'cut and-paste' strategy that children follow in language acquisition" by finding precedents in previous speech on the basis of idealised operations.

All methodological considerations reflected on here are also highly relevant to the current study. I will only look at utterances that include spatial language and regard both parental input and child speech as my main corpus because I am interested in the link between linguistic environment and language acquisition. I will further use the three operations SUPERIMPOSE, SUBSTITUTE and ADD proposed in Koch (2019) to derive target utterances from their closest matches. In that sense, following Hartmann et al. (2021, pp. 240-243), my traceback study, too, will be both "confirmatory" and "explorative". On the one hand, it aims at corroborating the usage-based model of language acquisition and, on the other hand, it intends to impart further information on the linguistic representation of space in early childhood.

4.3. Study Design

I shortly map out my own procedure, which, in fact, involved several decisions and preparatory measures prior to conducting actual tracebacks. Figure 8 illustrates the steps taken during my project; the grey boxes represent the respective stages whereas the green arrows indicate sequence:

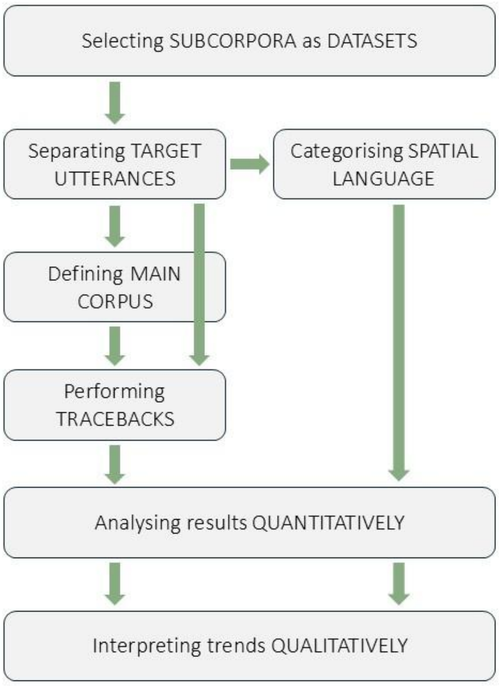


Figure 8. Visualisation of the steps taken in the present study

As suggested in the literature, the first requirement in any successful corpus-based language acquisition study is to choose a suitable dataset (Demuth, 2008, p. 200; Corrigan, 2011, pp. 274-275). Demuth (2008, p. 200) mentions just a few of the crucial considerations on which corpora should be created:

> the number of children to be included in the study, the setting for recording (home, lab, school), the interlocutors (parent, siblings, experimenter), the activities ('natural', prompted with a specific set of toys/tasks), the amount of data recorded (how long recordings should be) [and] the number of sessions/ages recorded per child (i.e., longitudinal or not, how frequently sampled).

I have already described the speech samples I used and called attention to the criteria by which I designed my three subcorpora in 4.1. After determining a test corpus for each subject aged 2;6, I carried out 2 strands in parallel to bring together in the analysis and discussion sections: one concerned spatial language categories (arrow to the right) while the other entailed selecting the target utterances to be traced back at a later stage of this study (down). Being interested only in children's spatial language, I disregarded all productions by caregivers as well as constructions without any reference to space. Furthermore, metaphorical extensions of the spatial domain, like *in a minute* (Eleanor 020600a, 929), and uses of the future tense, *I gonna [: going to] read it* (Fraser 020601a, 399), for example, were not included either.

I then assigned all genuine spatial terms as produced by the children to one of six groups following the typology suggested by Ferrara et al. (2011) and Yang and Pan (2021). This served the purpose of, one the one hand, demonstrating the significance of linguistic representations of space in early multiword speech and, on the other, establishing a basis to ground my derivations. In sum, I categorised between 165 (Fraser) and 180 spatial language utterances (Eleanor) per child. I recorded the coding in Excel datasheets.

On this basis, I further identified constructions that were not eligible for the traceback approach (see Hummel, 2006, pp. 2-3). Reasons for exclusion included if the strings were unintelligible (marked as xxx, yyy or www in the transcript; MacWhinney, 2024 for a comprehensive summary of the CHAT conventions), erroneous (*), imitations or self-repeats (+I, +SR). Once all target utterances that lend themselves to reconstructing were identified, I familiarised myself with the materials and defined three main corpora of about the same size in which I would look for precedents (section 4.1). Hummel (2006, p. 4) in her traceback handbook distinguishes three types of matches: exact matches, created matches and failed derivations.

Exact matches are fixed constructions that have been said in their entirety before; one can presuppose that the children have stored a mental representation of such productions since they do not need to alter known material to create them (Koch, 2019, pp. 180-181). In order to be considered a zero-operation traceback, the target utterance has to occur at least twice in the main corpus (Dąbrowska & Lieven, 2005, p. 447). If, however, a string is not a direct repeat, investigators attempt to reconstruct it by adding, substituting or superimposing component units (Koch, 2019, p. 181). In those cases, children use schemas with open slots, “expression[s] for which there are at least two previous examples of shared lexical material together with variation in the same position” (Hummel, 2006, p. 5), to create novel utterances. Lieven et al. (2009, pp. 486, 489) framed the rules that researchers should select the longest possible schemas and fill slots were with the longest available units of the suitable semantic category to arrive at a match. The traceback handbook lists some illuminating sample sentences to demonstrate permissible derivations (Hummel, 2006, p. 10):

(12) target utterance: *I got my head.*
head schema: *I got my REF.*
precedents: *I got my socks on*. ✗
I got my own. ✓
I got my bag. ✓

(13) target utterance: *Katie drink her milk*.
precedents for *drink*: *I want a drink.* ✗
I wanna drink juice. ✓

Example (12) illustrates that the longest schema, *I got my REF*, not *I got REF*, is used. Moreover, the head schema must always frame the entire utterance; *got on* with its additional preposition in the final position does not fit the pattern and is thus not considered in the analysis. The noun *drink* in (13) further cannot count as a precedent since it does not share the syntactic and semantic sense of the verb *drink* in the target construction (Hummel, 2006, pp. 10, 12).

Whenever several paths of reconstructing were possible, the one requiring fewer operations was chosen, assuming that children would opt for the easiest route of deriving utterances (Koch, 2019, p. 183). In addition, Dąbrowska and Lieven (2005, p. 460) assert that SUPERIMPOSE has priority over SUBSTITUTE as partial lexical overlap would facilitate assembling chunks, just "like putting together two jigsaw puzzle pieces which have pictures of a part of the adjoining piece attached to them". Operations of substitution, in turn, override juxtapositions. In juxtapositions, material can be added both at the beginning and end of utterances and consist of more than one part (Hummel, 2006, p. 12).

Where previous studies used computerised programmes unavailable to me, like The Autotracer (e.g. Lieven et al., 2009) or Tracebacker (e.g. Koch, 2019), I relied on CHILDES's own software CLAN (MacWhinney, 2025) to manually search for the links between target utterances and prior strings. I downloaded the audio files of the main corpora onto my computer and selected them as my directory; this was done separately for each child. Hummels (2006, pp. 30-31) suggests running the commands KWAL and COMBO to find individual words or larger patterns in the main corpus; examples are given in the following table:

Table 9. Description of the commands used in CLAN

KWAL	COMBO
Searches for a specific word or group of words (MacWhinney, 1996, p. 10)	Searches for combinations of words or types of words (MacWhinney, 1996, p. 10)
`kwal +s“word” file.cha`	`combo +s“word1^word2” file.cha`
> kwal +sdog *.cha UD features file being used: "C:\talkbank\clan\lib\fixes\UD_features.cut" kwal +sdog *.cha Tue Jan 28 14:49:45 2025 kwal (09-May-2024) is conducting analyses on: ALL speaker tiers ** From file <020400a.cha> From file <020400b.cha> -- *** File "020400b.cha": line 2819. Keyword: dog *MOT: oh (.) has your dog come to xxx ? [+ PI] • -- *** File "020400b.cha": line 2847. Keyword: dog *CHI: dog [?] , you do it . • From file <020401a.cha> -- *** File "020401a.cha": line 166. Keyword: dog *CHI: dog ! • -- *** File "020401a.cha": line 170. Keyword: dog *INV: it's a dog . • -- *** File "020401a.cha": line 173. Keyword: dog *INV: <what does a> [/] what does a dog say ? • -- *** File "020401a.cha": line 176. Keyword: dog *CHI: <I'm a dog> [?] . • -- *** File "020401a.cha": line 837. Keyword: dog *INV: I'm gonna [: going to] give them to my dog to eat . • --	> combo +s"this^is" *.cha UD features file being used: "C:\talkbank\clan\lib\fixes\UD_features.cut" this^is combo +sthis^is *.cha Tue Jan 28 14:46:49 2025 combo (09-May-2024) is conducting analyses on: ALL speaker tiers ** From file <020401.cha> -- *** File "020401.cha": line 3559. *MOT: do you know who (1)this (1)is , Fraser ? • -- *** File "020401.cha": line 4686. *MOT: (1)this (1)is another story here . • Strings matched 2 times From file <020402a.cha> -- *** File "020402a.cha": line 496. *MOT: &-er (1)this (1)is the freezer . • -- *** File "020402a.cha": line 2106. *MOT: (1)this (1)is your lunch here . • Strings matched 2 times From file <020402b.cha> -- *** File "020402b.cha": line 1484. *MOT: (1)this (1)is your blanket . • Strings matched 1 times

In line with the formulae listed above, I entered queries into CLAN’s command window to detect particular strings. The commands KWAL or COMBO were followed by a mandatory `+s` that specifies the linguistic unit and the unit itself in quotation marks, for instance, `+s“dog”`. If I searched for combinations of words, larger constructions or full utterances, a circumflex symbol separated individual units, like in `+s“this^is”`. The last position of the query states which recordings the system ought to screen; one can either give the exact number of single files, `020601a.cha`, or, as I predominantly did, use the Boolean operator * to browse all provided files in one go. The asterisk was, furthermore, helpful in running less specific queries; it was inserted to locate pattern with slots in the middle, such as `“I^*^it”`, lemmas, `“car*”` for car, cars, car’s and cars’, or different words of the same form, including `“*ed”` and `“*ing”`. The programme lists all hits in the main window and indicates in which file they were detected (see Hummel, 2006, pp. 30-32 for recommendations).

After documenting the results and some additional information on the traceback, i.e. derivation strategies, head schemas and component units, for each of the children's target utterances in an Excel file (see Appendix for extract), I synthesised the two strands of research. Moreover, I qualitatively evaluated and interpreted them against the background of previous traceback studies and spatial language learning literature.

Thus far this thesis has surveyed literature on the acquisition of multiword sequences and spatial terms and presented the method and procedure used in this current investigation. In the section that follows I move on to provide the principal findings obtained and make some informed statements about spatial terms in multiword speech and usage-based language acquisition.

5. Analysis and Findings

Turning now to the results of this thesis, I first give evidence illustrating that children convey a plurality of spatial information already at an early stage of language development. Subsequently, I discuss the main outcomes of my traceback study, highlighting successful matches, operations and slot-and-frame patterns but also giving possible reasons as to why some reconstructions failed.

5.1. Spatial Language

It is known from previous research that (a) spatial vocabulary emerges at a young age and that (b) talk about the spatial world impacts children's cognitive development. To further contribute to the debate, I analysed data from the test corpora, consisting of approximately 1 hour of interaction between child and their caretaker, largely the mother, which allows me to draw conclusions about the naturalistic use of language of space. These recordings amount to 1631 produced utterances on average across all interlocutors; between 38.8 and 43.8 percent of the conversations are child-led. In total, 180 (Eleanor), 165 (Fraser) and 173 constructions (Thomas) are of vital importance to this analysis as they contain at least one linguistic representation of space. Some of those strings exhibit more than one spatial language token, resulting in 647 constructions after adding up the numbers for the three children (Ø = 215.7, SD = 12.8). The overall findings are displayed in Table 10:

Table 10. Frequency of spatial language tokens in absolute numbers

	Eleanor 020600a.cha	Fraser 020601a.cha	Thomas 020602.cha
Number of utterances			
Child	668	830	557
Parent	858	1101	878 (*MOT: 871 + *FAT: 7)
Total	1526	1931	1435
Number of child utterances containing spatial terms	180	165	173
Spatial language tokens	221	198	228
Spatial locations	82	56	69
Deictic terms	95	96	37
Dimensions	10	13	22
Shapes	2	1	18
Spatial orientations or transformations	32	31	81
Spatial features or properties	1	1	1

As can be seen from the last row of data, there is exceptionally good agreement between the distribution of spatial language for Eleanor and Fraser at age 2;6. Deixis is represented best (95x Eleanor, 97x Fraser), followed by spatial locations as the second strongest group (82x Eleanor, 56x Fraser). What is surprising, however, is that both of these categories are trumped significantly in Thomas's corpus by spatial orientations and transformations; Thomas uses more than twice as many compared to 32 or 31 tokens in Eleanor's and Fraser's datasets. Another interesting result to emerge from this initial analysis is that the quantity of spatial locations is larger in the Thomas-corpus (69x) than in the Fraser-corpus (56x) but does not quite reach absolute numbers as high as in the Eleanor-corpus (82x). At the same time, for every two to three deictic expressions produced by Eleanor or Fraser, Thomas utters just one. Despite these evidently stark divergences in pointing to places or situations and describing dynamic scenes, all three children display similar linguistic behaviour with regards to the other categories, such as dimensions (10x Eleanor, 13x Fraser, 22x Thomas) and spatial features and properties (1x Eleanor, 1x Fraser, 1x Thomas). Moreover, shapes are fairly consistent across the corpora as well with only 1 to 2 types represented in the datsets, or 1 to 18 uses. This chart illustrates the relative distribution of spatial terms, making visible similarities and differences:

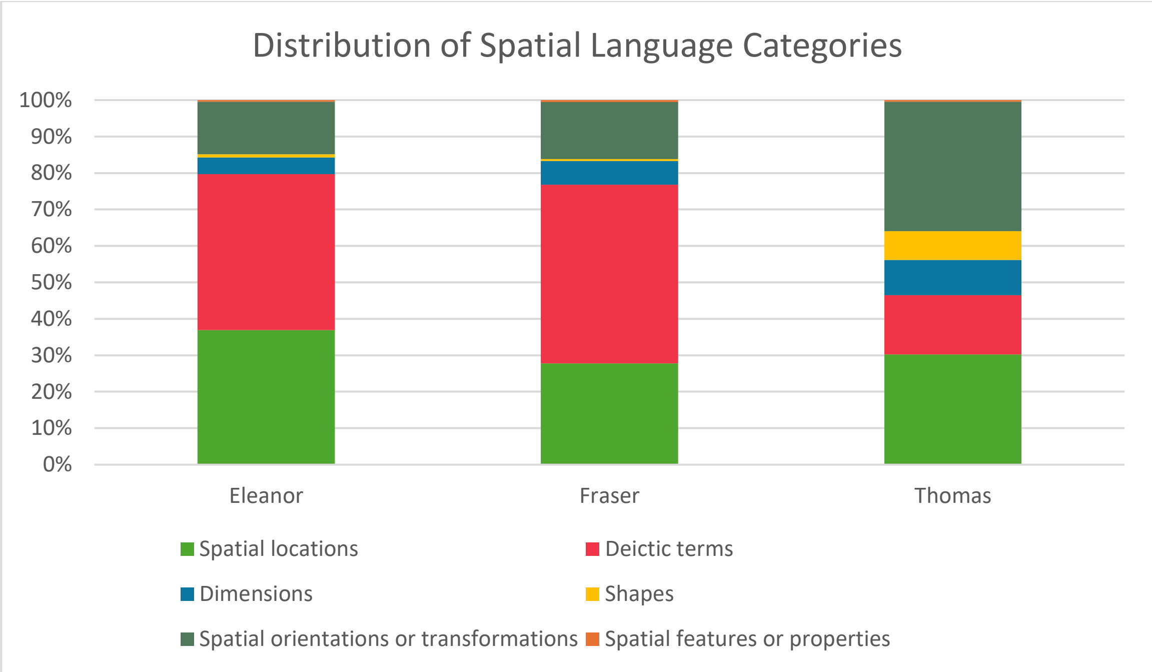


Figure 9. Distribution of spatial language in Eleanor, Fraser and Thomas's test corpora

A clear picture of children's spatial language inventories at 2;6 emerges when examining only Eleanor and Fraser's column in Figure 9, yet Thomas's low use of deictic terms and heavy reliance on orientations and transformations is slightly puzzling. I will now discuss all six categories, attempting to find explanations for the patterns I observed in the data.

First, spatial locations appear to be of equal relevance to all three children's linguistic references to space. Fraser and Thomas produce terms denoting spatial locations in 3 out of 10 cases while Eleanor's share is slightly higher at 37.1 percent. In general, spatial locations were very often specified with dimensional prepositions and prepositional phrases. The children situate trajectors *at* 0-dimensional points, *it's a boy at the wood* (Eleanor 020600a, 1319) and *look at this book* (Fraser 020601a, 1169), *on* 2-dimensional surfaces, *on the grass* (Fraser 020601a, 1212), *on it* (Eleanor 020601a, 294 & 718) and *on the side* (Thomas 020602, 176), or *in* 3-dimensional containers, *racing car going in there* (Thomas 020602, 410) and *she's in the pushchair* (Eleanor 020600a, 783). Beyond the abundance of locative uses, Eleanor and Thomas also infrequently utter directions, describing goals, *she [*] gone up to bed* (Eleanor 020600a, 1126), as well as sources, *from home* (Thomas 020602, 1292). Thomas, who is very much still transitioning from telegraphic to multiword speech, occasionally omits function words, producing *no up* and *no people train* (Thomas 020602, 138 & 622) to mean 'don't go up there, policemen' and 'there are no people on the train', respectively. The last example further demonstrates that the learners also conceptualise concrete nouns as landmarks, *farmyard* (Eleanor 020600a, 380) or *Torremelenos [sic]* (Thomas 020602, 31), to name a few. All three children also have in common the use of interrogative pronoun *where* to ask for the location of entities and situations (16x Eleanor, 8x Fraser, 2x Thomas). The data thus implies a certain urgency for children to communicate spatial locations relative to the position they themselves occupy in order to organise their immediate environment and create shared intentionality and joint action with their caregivers.

Another aspect of spatial language that stands out in the corpora is the occurrence of deictics to delimit spatial scope. Deixis I observed consists above all of spatial adverbs, *here* and *there*, and demonstratives, *this*, *that*, *those*, *this one*, *that one* or *this space*. Close to half of Eleanor's and Fraser's spatial constructions contain such deictic terms. While spatial deixis is regularly used to point to specific trajectors, *that mouse* (Fraser 020601a, 244), or places, *sit there* (Eleanor 020600a, 560), it is also at times embedded in more complex structures, for one thing, prepositional phrases, *Mummy goes to there* (Eleanor 020600a, 475) and *is [*] giraffe in here?* (Fraser 020601a, 588). Thomas, a less precocious speaker, on the other hand, focuses on content words and misses grammatical elements on occasion, as can be seen in *big tunnel* and *train come* (Thomas 020602, 595 & 763). Sentences like these indicate that he still views spatial adjuncts and demonstrative pronouns as optional, using them rather in set phrases, *what's this?*, *Daddy's here* and *that one* (Thomas 020602, 257, 1181 & 1228), for example, than systematically. This explains the significant gap in proportion of deictics between the three children's datasets, too. Together these results suggest that learners already divide space and know to shift the view of the interlocutor through linguistic devices.

With percentages from 4.5 to 9.6, dimensions do not feature as strongly as locations and deictic terms. Children represent variations in size, from *little* and *small* to *big*, *deep* and *high*, by producing select adjectives. Attributive uses, *all the big books* and *I see a red big dragon* (Fraser 020601a, 95 & 656), heavily outweigh predicative ones, *are you just tall, Polly?* (Eleanor 020600a, 233). The strings *really high* (Fraser 020601a, 1509), *too big* (Thomas 020602, 772) or *little bit big* (Eleanor 020600a, 1137) show that dimensions are also modified, although rarely. Examining the nouns following words denoting dimensions further reveals that they predominantly occur in combination with semantic fields of body parts, *my big stomach* and *a small stomach* (Eleanor 020600a, 51 & 62), expressions of affection, *big squeeze*, *big kiss* and *big hug* (Fraser 020601a, 795, 800 & 806) and transportation, *big bus* and *big tunnel* (Thomas 020602, 26 & 599).

Both shapes and spatial features and properties are uncommon structures, occurring just 1 to 2 times per child. While Eleanor (020600a, 1311 & 1356) talks of *a piece* and *a yellow piece*, Fraser (020601a, 1148) uses the sentence *I'm building a tower*. Thomas's productions differ from those of the other two children by producing 18 spatial terms (7.9 percent) classified as shapes. Upon closer inspection, however, this outlier can be accounted for: Many self-repeats in succession give the impression that Thomas productively uses shapes, when it is, in fact, just two types, *tunnel* (9x) and *box* (9x), and hence well within the same range as Eleanor and Fraser. For each child only 1 in 200 tokens in the corpora was coded as a spatial feature or property; the sole instances include *upstanding* (Eleanor 020600a, 671), *upside-down* (Fraser 020601a, 1881) and *opened* (Thomas 020602, 844). One might thus argue that children in their third year of life prioritise information about locations over expressing spatial or geometric configurations. They ostensibly deem trajectors' positions in space more relevant to their interactions and do not necessarily become set on what they look like.

Lastly, the category of spatial orientations and transformations, the only one belonging to kinesis, is characterised a great degree of lexical variety, as example sentences (14), (15) and (16) from the children's datasets emphasise:

(14) How do we fold it? (Eleanor 020600a, 115)

(15) I wanna open the door. (Fraser 020601a, 887)

(16) Oh dear lid come off. (Thomas 020602, 38)

The motion verbs *fold*, *open* and *come off* stress that young learners successfully observe and characterise positional changes of entities and discern modes as well as directions of movement. Spatial orientations and transformations account for 15 percent of Eleanor and Fraser's spatial language tokens whereas Thomas shows a considerably higher share with a third of his constructions assigned to the group of orientations and transformations. Again, this disproportionally high number seems to be caused by the fact that he often conservatively repeats utterances, for example, *back fall down* (11x), *go(ing) in* (7x) or *fit* (5x) within a short time. In sum, I observe some manner-of-motion verbs, such as *slide*, *fly*, *ride* and *walk*, but for the most part path verbs, like *come* or *get out* on the horizontal axis as well as *fall*, *jump*, *climb* and *drop* to denote vertical motion. Children evidently use a great diversity of orientational structures, nonetheless a tendency towards static, as opposed to dynamic, representations of space is evident.

In this section I have compared the children's spatial language manoeuvres to find that, on the whole, they compare well. Spatial locations and deictic terms occur especially frequent, whereas shapes and spatial features are outnumbered significantly. Thomas displays an apparent preference for dynamic spatial expressions, that is, orientations and transformations, which I have traced to the telegraphic and repetitive nature of his speech. I have set up these initial observations for the traceback study to follow.

5.2. Successful Tracebacks

The present study was designed to determine the degree of interconnectedness in early multiword speech and to highlight the lexical and syntactic changes necessary for a child to derive novel productions from prior ones. Following the steps detailed in the methodology section, it turned out that a strong majority of target utterances containing spatial terms, but not all of them, can be reconstructed from precedents in a delineated main corpus. The three pie charts in Figure 10 stress that between 7 and 8 in 10 child productions can successfully be accounted for when tracing them back to earlier constructions:

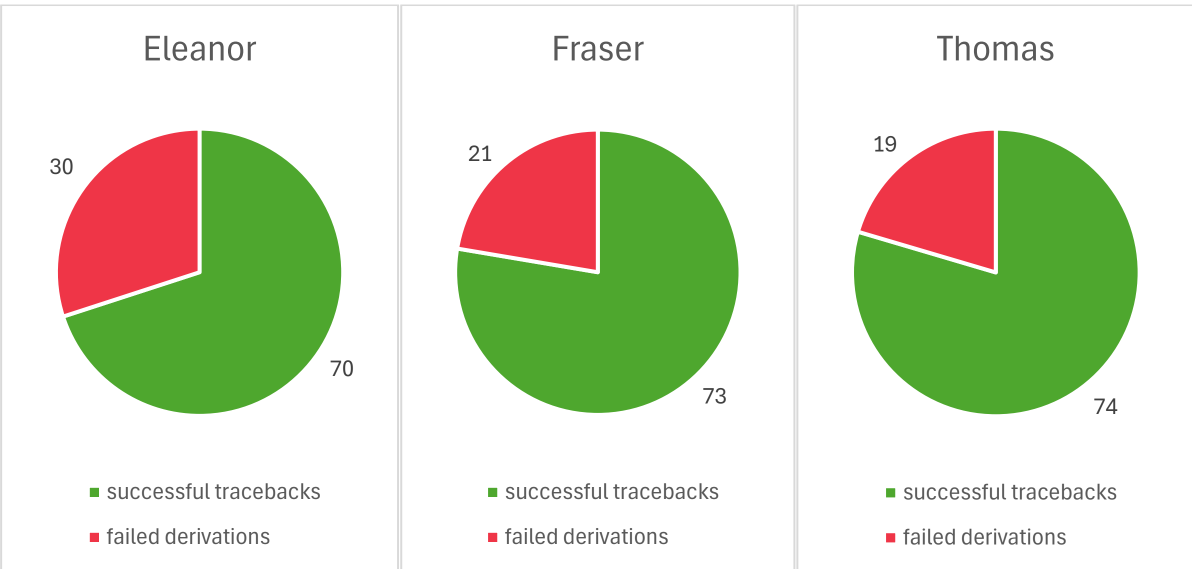


Figure 10. Proportion of successful tracebacks for Eleanor, Fraser and Thomas

For Eleanor, the lowest percentage of utterances, 70 percent to be exact, was derivable while matches could be created for 77.7 percent in the Fraser-corpus and 79.6 percent in the Thomas-corpus. These successfully reconstructed strings are allocated to four categories: exact matches, 1-operation tracebacks, 2-operation tracebacks and tracebacks requiring three or more operations, as summarised below:

Table 11. Distribution of operations in the corpora

Number	Operation	Eleanor		Fraser		Thomas	
		#	%	#	%	#	%
0		16	22.9	23	31.5	25	33.8
1	SUBSTITUTE	19	27.1	18	24.7	28	37.8
	SUPERIMPOSE	7	10.0	13	17.8	5	6.8
	ADD	1	1.4	0	0	5	6.8
	total	27	38.6	31	42.5	38	51.4
2		20	28.6	16	21.9	11	14.9
3 or more		7	10.0	3	4.1	0	0

* rounded to one decimal place

Data in Table 11 demonstrates that between one fifth and one third of the constructions has been said in their entirety before, either by the child, mother or any other interlocutor represented in the corpus. The largest share, ranging from 38.6 (Eleanor) and 51.4 percent (Thomas), in each child's dataset is formed by utterances requiring one operation. Considerably fewer productions can be derived by applying two (28.6% Eleanor, 21.9% Fraser, 14.9% Thomas) and three or more operations (10.0% Eleanor, 4.1% Fraser, 0.0% Thomas). It is evident that a substantial number of children's early multiword strings can, in fact, be accounted for using lexically specific units; however I also found instances of non-creative language use.

Zero-operation tracebacks, 16 to 25 utterances per child corpus, are available to the child because they have been produced before; all of them have been itemised in the following table:

Table 12. Exact matches in the corpora

Eleanor 020600a (N=16)	Fraser 020601a (N=23)	Thomas 020602 (N=25)
CHI: oh there you are. (14)	CHI: I dropped it. (78)	CHI: like that. (73)
CHI: in there. (16)	CHI: do like this. (329)	CHI: none there. (104)
CHI: there. (142)	CHI: I want this one. (583)	CHI: no up. (138)
CHI: jump. (172)	CHI: there's one! (612)	CHI: that's better. (218)
CHI: where [>]? (281)	CHI: that's you. (715)	CHI: what's this? (257)
CHI: push it. (323)	CHI: and a big hug. (806)	CHI: open it. (391)
CHI: can I go? (503)	CHI: it's dark in here. (900)	CHI: going backwards. (418)

CHI: sit there. (560)	CHI: that's a funny noise. (925)	CHI: box. (525)
CHI: here. (638)	CHI: where's Daddy? (1018)	CHI: fit in there. (534)
CHI: here you go. (666)	CHI: where's Jwww [% brother]? (1022)	CHI: fit in. (535)
CHI: open your mouth. (798)	CHI: downstairs. (1026)	CHI: fit. (548)
CHI: <get out>. (852)	CHI: I'm building a tower. (1148)	CHI: coming from. (571)
CHI: let’s go. (1077)	CHI: on [?] the grass. (1212)	CHI: coming. (609)
CHI: that one. (1185)	CHI: open the gate. (1317)	CHI: coming in there. (614)
CHI: little (?) one. (1277)	CHI: do that. (1331)	CHI: moving. (662)
CHI: go away. (1493)	CHI: there! (1708)	CHI: too big. (772)
	CHI: there it is! (1709)	CHI: what's this? (851)
	CHI: what's that? (1768)	CHI: out_of the way. (913)
	CHI: and this? (1770)	CHI: that one. (967)
	CHI: what's this? (1772)	CHI: here you are, Mummy. (970)
	CHI: and what's this? (1775)	CHI: &~de supermarket. (1066)
	CHI: but what's this? (1854)	CHI: <Mummy here> [?]. (1199)
	CHI: what's this one? (1871)	CHI: &~isa back. (1232)
		CHI: what in there? (1249)
		CHI: there. (1252)
		CHI: big willie. (1381)

The general assumption is that these strings occur frequently in the children’s linguistic environment so that the become entrenched and are rote-learned as a whole. In comparing the exact repeats, similarities between Eleanor, Fraser and Thomas become evident. First, some of the fixed strings, for instance, *here you go* and *here you are, Mummy*, which accompany the act of passing entities to a recipient, or *and a big hug*, which is always uttered when Fraser and his mother embrace, are established linguistic routines. They are arguably linked to certain social interactions and possibly activated as units, thus facilitating the navigation of these contexts for the child. Preconstructed formulae, moreover, especially often take the shape of interrogatives, including questions like *but what’s this*, *where’s Daddy?* and *what in there?*, or imperatives, commands such as *jump*, *let’s go*, *push it*, *open the gate* or *do that*. They might be easier to produce from memory due to the fact that their structure is distinct from declarative sentences. Other productions can be interpreted as commentary on activities performed during the recording sessions; as the conversations are always set during free play, the topics talked about are rather limited. Much of the child’s speech, therefore, relates to snacks (*I dropped it*), toys (*I’m building a tower*) or entities’ movements in space (*going backwards*, *fit in*). The frozen units are generally all short and feature several deictic terms, *this*, *that*, *here* and *there*, for example. I have already mentioned the importance of deictics in early spatial language; I assume that they occur as set phrases more frequently than spatial locations, *the supermarket* or *on the grass*, since deixis is not bound to specific entities but can be used to refer to various ones. Due to this applicability to many situations, and subsequent greater use, preconstructed chunks containing deictic terms might be acquired faster.

Much more interesting to this project, however, are constructions which are less repetitive and partially productive. To reiterate, the traceback approach postulates that children apply the three operations SUBSTITUTE, SUPERIMPOSE and ADD to build novel constructions from the main corpus. The target utterance *a big bus* was successfully derived from previous constructions found in Thomas's corpus and required only one operation to match precedents:

Table 13. Tracing back *a big bus* (Thomas 020602, 26)

Component units	Operations	Strategy	Attested frames or fixed strings	Attested slot fillers
a big [REF] bus	1	SUBSTITUTE	MOT: have I got a big tummy? CHI: a big noise. CHI: a big playgroup.	CHI: blue bus. MOT: on the bus? CHI: lady ride the bus.
Successful traceback: *a big bus*				

The slot-and-frame pattern *a big [REF]* is highly productive in the dataset, occurring 175 times with a diverse set of referents inserted into the open slot. The plentiful instantiations of this schema ostensibly allow the child to notice regularities, *a big* is always followed by a noun, *tummy*, *noise* or *playgroup*, for example, and abstract across them. This leads to the production of the schematic construction *a big bus* where *bus* is substituted into the REFERENT slot. There is ample evidence that Thomas knows the component unit *bus* from other linguistic cotexts, such as *blue bus*, *on the bus* and *lady ride the bus*.

Another example for the SUBSTITUTE operation, from Eleanor's dataset, is given in Table 14. The pattern *where's [REF]?* features frequently enough to leave a cognitive trace and thus be extended to the formation of the novel constructions *where's Jeannine?*:

Table 14. Tracing back *where's Jeannine?* (Eleanor 020600a, 107)

Component units	Operations	Strategy	Attested frames or fixed strings	Attested slot fillers
where's [REF]? Jeannine	1	SUBSTITUTE	MOT: where's your dog? CHI: where's the ticket? CHI: where's Daddy?	MOT: Jeannine's asking you a question, darling. CHI: Jeannine. CHI: can I play with that Jeannine?
Successful traceback: *where's Jeannine?*				

Both tracebacks have in common that it is not the spatial term, *big* or *where*, that is substituted but a referent in the final slot. In *put it on the radiator*, too, the open slot is not the LOCATION slot, *on the radiator*, but just the REFERENT, *the radiator*, because the traceback asks for the longest available lexically specific schema.

In contrast to Tables 13 and 14, there is overlap between the slot filler and the fixed constituents in the schema *put it on [REF]*. The missing component unit, *(on) the radiator*, is inserted using the SUPERIMPOSE-operation, as the *on* of the component unit and *on* of the schema coincide in the target utterance *put it on the radiator*:

Table 15. Tracing back *put it on the radiator* (Eleanor 020600a, 302)

Component units	Operations	Strategy	Attested frames or fixed strings	Attested slot fillers
put it on [REF] (on) the radiator	1	SUPER-IMPOSE	FAT: put it on the wall. MOT: don't put it on the floor. CHI: put it on my foot.	MOT: do you wanna put your head down on the radiator?
Successful traceback: *put it on the radiator*				

Just like *put it on the radiator*, the string *I can open it* in the Fraser-corpus is also created by through superimposition since the lexical element *it* is part of both the component unit *open (it)*, which fills the PROCESS slot, and the item-based schema *I can [PRO] it*.

Table 16. Tracing back *I can open it* (Fraser 020601a, 894)

Component units	Operations	Strategy	Attested frames or fixed strings	Attested slot fillers
I can [PRO] it open (it)	1	SUPER-IMPOSE	CHI: I can have it. CHI: yeah I can see it. CHI: I can do it! CHI: I can carry it.	INV: do you wanna open it? MOT: you open it now then. CHI: open it for me.
Successful traceback: *I can open it*				

As can be deduced from this table, slot fillers attested in the dataset are not semantically restricted to spatial orientation verbs like *open* but also cover *have*, *see* and *do*, to name a few. The frame is thus not inherently spatial in nature but allows words of all kinds of meanings, provided they fit the word class of the open PROCESS slot. In fact, the frame *I can [PRO] it* can be realised as a linguistic representation of space should a dynamic spatial term, for example, *open*, be inserted.

For all the target utterances discussed thus far to be built from previous utterances in the corpora, the children applied the operations SUBSTITUTE and SUPERIMPOSE. Some tracebacks, however, did not require partially productive slot-and-frame pattern but were formed through juxtaposing lexical material. The ADD operation affects instances in which children string together a fixed string and an extra constituent, such as *this one* and *Mummy* in the example below:

Table 17. Tracing back *this one, Mummy* (Thomas 020602, 966)

Component units	Operations	Strategy	Attested frames or fixed strings	Attested slot fillers
this one + Mummy	1	ADD	CHI: this one.	CHI: telephone now [?], Mummy.
Successful traceback: *this one, Mummy*				

The fixed string *this one* occurs in 44 constructions in the Thomas-corpus; 20 times it is produced in isolation, constituting a full utterance. In the target utterance at hand, *this one* is modified by adding the vocative *Mummy*. Across all three datasets, vocatives, especially *Mummy*, are commonly attached to frozen phrases, so are the conjunctions *and* and *but* as well as *because*. Fixed formulae also combine with polite expressions, *please* and *thank you*, interjections, *yeah* and *oh dear*, and, most interestingly, adjuncts of space and time, including *(in) here*, *(in) there* and *now*. This suggests that children use the juxtapositions of deictic spatial adverbs as a strategy to specify the location of referents and anchor events in space.

Figure 11 shows that the ADD operation is in fact used very infrequently in 1-operation tracebacks compared to SUBSTITUTE and SUPERIMPOSE, maximally in 13.2 percent of constructions:

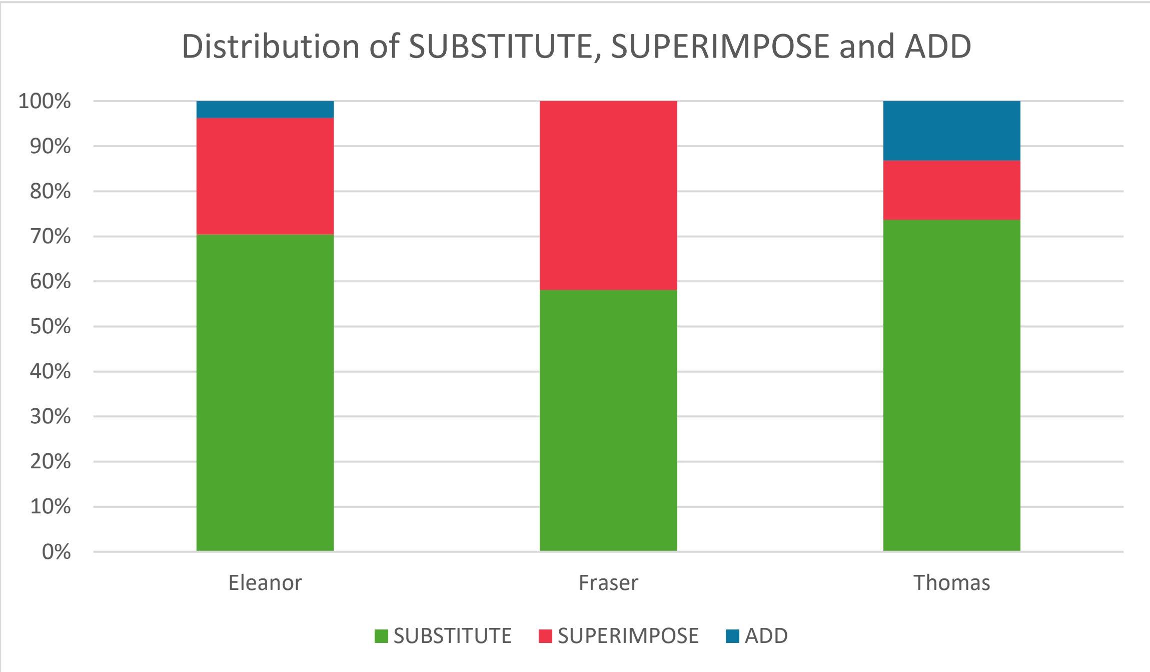


Figure 11. Use of operations in utterances requiring one modification

Eleanor builds novel utterances by juxtaposing lexical material once (3.7%), whereas Fraser only substitutes and superimposes but never adds strings from earlier productions and input. By implication, for most constructions requiring just one alteration to match precedents, the operations SUBSTITUTE and SUPERIMPOSE play a key part. Even across all tracebacks, considering also multi-operations, the children use ADD only between 9 (Fraser) and 12 times (Eleanor). This underlines that early construction learning is indeed preponderantly based on low-level schemas with open slots, with at least 86.8 percent of successful single-operation derivations being lexically specific in nature.

Yet some constructions in the test corpus exhibit an even higher degree of creativity, requiring more than one modification to match precedents. 2-operation tracebacks amount to 14.9 percent of successful derivations for Thomas, 21.5 percent for Fraser and 28.6 percent for Eleanor. Two examples illustrate that such productions are created either on the basis of a frequent schema with two open slots (Table 18) or through adding a lexeme to a pattern with one variable position (Table 19):

Table 18. Tracing back *is that Liz [: Liz's] [*] bedroom?* (Fraser 020601a, 1687)

Component units	Operations	Strategy	Attested frames or fixed strings	Attested slot fillers
is that [POSS] [REF]? Liz's bedroom	2	SUBSTITUTE SUBSTITUTE	CHI: is that Lala's ball? CHI: is that my balloon? MOT: is that Mummy's plate?	MOT: is Liz [%aunt] upstairs? CHI: I needta go to Jwww's [%brother] bedroom. MOT: let's go to your bedroom.
Successful traceback: *is that Liz [Liz's] [*] bedroom?*				

Table 19. Tracing back *not get in there* (Thomas 020602, 396)

Component units	Operations	Strategy	Attested frames or fixed strings	Attested slot fillers
not [PRO] get in + there	2	SUBSTITUTE ADD	CHI: not see. CHI: not cry.	CHI: sit in there. MOT: get in your chair please.
Successful traceback: *not get in there*				

The tables univocally show that the first utterance is less specific than the latter, displaying a higher degree of abstraction with two variable slots, POSSESSOR and REFERENT. The *is that [POSS] [REF]?* slot-and-frame pattern is well documented in Fraser's corpus with exemplars such as *is that Lala's ball?* or *is that Mummy's plate?*. He generalises from those prior utterances that the first slot semantically has to be filled with a word referring to a possessor, e.g. possessive pronouns, *my*, or proper names, *Mummy's* and *Lala's*. The second slot is fitted with things possessed, like *ball*, *balloon* or *plate*, and must, by analogy, also take the referent *bedroom*. The target utterance *is that Liz's bedroom?* can thus be reconstructed by applying the SUBSTITUTE operation twice.

The sentence in Table 19, on the other hand, is more item-based as it only permits lexical alterations in one position, the PROCESS slot. Based on the attested constructions in the main corpus, I argue that the schema *not [PRO]* is already established and can be added to. With the SUBSTITUTE operation, Thomas inserts *get in* into the open slot; in order to arrive at a match, however, the utterance is crucially changed in a different regard as well, that is, the string *not* *get in* is then coupled with the adverb of space *there*. *Not get in there* is arguably accounted for by tracing back to the schema *not [PRO]* and attaching a spatial *there*.

Then again, other tracebacks require even more complex processes to build the target utterance from the corpus. Eleanor's production *and I see it there* is less closely related to precedents on a lexical level, sharing only *I* and *there* as well as the semantic restrictions of the open PROCESS and REFERENT slots with previously attested strings.

Table 20. Tracing back *and I see it there* (Eleanor 020602, 107)

Component units	Operations	Strategy	Attested frames or fixed strings	Attested slot fillers
and + I [PRO] [REF] there see (it) it	3	ADD SUPER- IMPOSE SUBSTITUTE	CHI: I have some dog food there. CHI: I can do it there like that. CHI: I will do it there.	FAT: I see you. MOT: I can see it. MOT: and I like cherry cakes.
Successful traceback: *and I see it there*				

The example in Table 20 depicts the interplay of all three operations, SUBSTITUTE, SUPERIMPOSE and ADD, in forming the novel utterance *and I see it there* from prior productions by the mother-child dyad. The production expands on the head schema *I [PRO] [REF] there* that also underlies *I will do it there* and *I have some food there*. Eleanor superimposes the chunk *see (it)* into the PROCESS slot, which incidentally overlaps with the *it* that is substituted into the ensuing REFERENT slot. There is ample evidence in the main corpus that the child has not only mastered the slot fillers *see (it)* and *it* but also knows that adding *and* at the beginning of an utterance is a linguistically valid operation. The traceback approach therefore equally allows piecing together constructions that exhibit a lower degree of specificity than the examples presented thus far.

Returning briefly to the numbers discussed at the beginning of section 5.5., it becomes evident that making three or more changes to utterances previously heard or produced is possible only in very few cases. Only a minority of successful tracebacks, 10 percent of Eleanor's and 4.1 percent of Fraser's constructions, can be linked to prior utterances by applying 3 or more operations. 4-operation tracebacks are applied particularly infrequently with only one occurrence in the Eleanor- and Fraser-corpus, respectively. Thomas, however, does not produce any construction for which 3 or more operations are required to create a match; it might well be that the lack of more advanced manoeuvres in the third child is a result of an early stage of language development.

As expected, the data shows a correlation between the number of operations Eleanor, Fraser and Thomas employ in constructing novel utterances and the children's mean length of utterance at 2;6. The following figure gives information about the proportion of derivations requiring no, one, two and three or more operations to arrive at a match for each child:

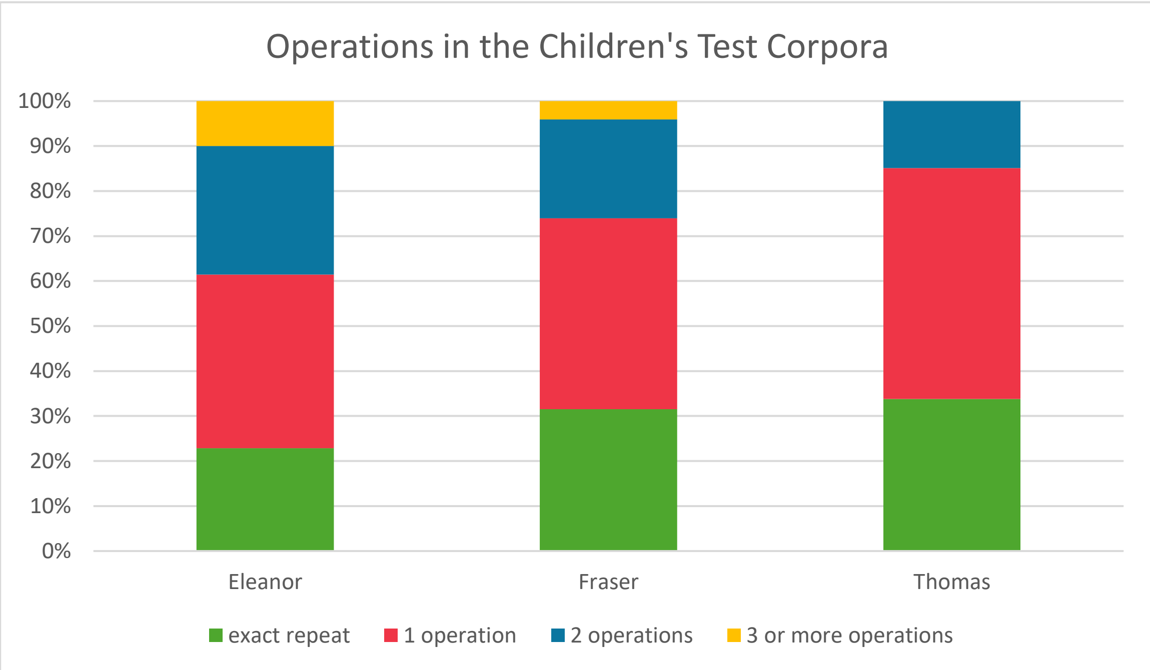


Figure 12. Exact repeats, 1- and multi-operation tracebacks among the target utterances

With respect to the MLU values calculated in section 4.1., it was found that Thomas, the child with the lowest MLU, uses the biggest share of exact repeats of previous utterances (33.8%), followed by Fraser (31.5%), then Eleanor (22.9%). His test-corpus also features the most tracebacks accounted for on the basis of just one modification, more than half as opposed to about 40 percent for Eleanor and Fraser. Taking a developmental perspective, this suggests that the share of zero-operation as well as 1-operation reconstructions falls steadily as the MLU increases. The reverse trend can be observed when it comes to target utterances created through multiple operations; at a mean length of utterance of 6.243, Eleanor achieves the highest number of successful tracebacks requiring two, three or more changes to be built form the main corpus (27 utterances; compare 19 for Fraser at an MLU of 5.480 and 11 for Thomas at an MLU of 3.400). One might conclude that while lower MLU tracebacks are more reliant on repetition and minor alteration, children produce more abstract and creative constructions with growing linguistic experience.

The analysis presented here has substantiated the claim that children at the early multiword stage of language acquisition depend on fixed formulae and lexically specific schemas when creating novel utterances. The strings that were traced back to CS or CDS attested in the main corpora vary in abstractness and number of operations required to match. I have, moreover, emphasised the close relationship between prior and newly formed child productions. Thus, examining the open slots in schematic constructions provides valuable insights into the linguistic material completing slot-and-frame patterns. The column chart below presents the slots filled by Eleanor, Fraser and Thomas:

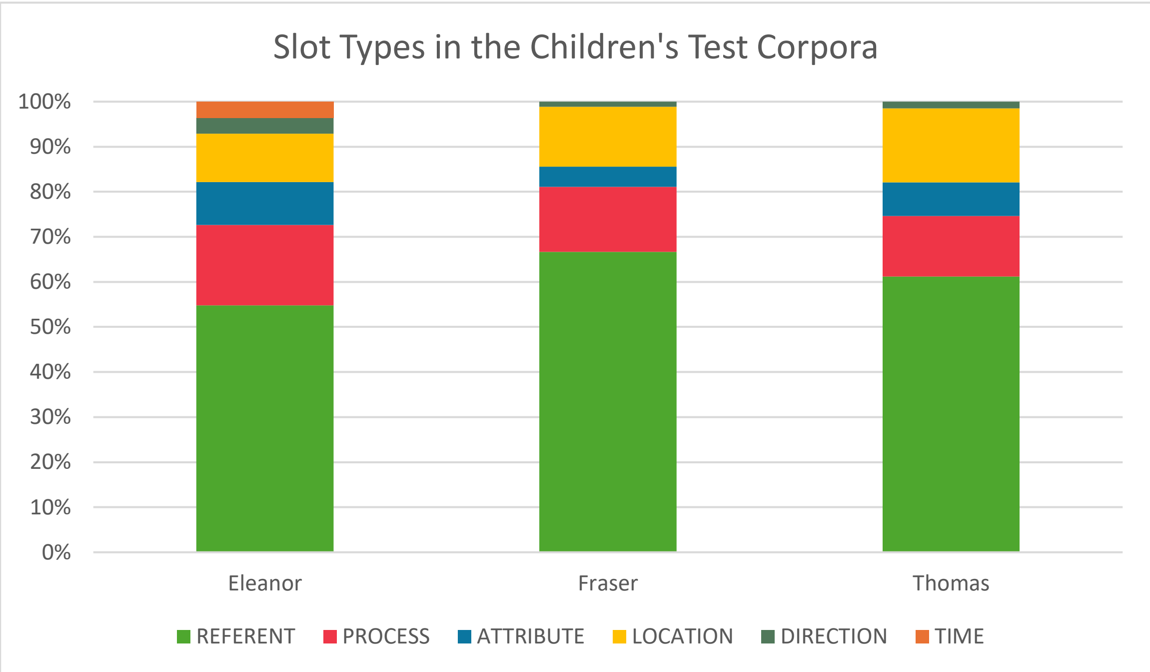


Figure 13. Percentage distribution of different slots in the datasets

Figure 13 emphasises that the overwhelming majority of open slots in successful tracebacks are of referents (54.1% Eleanor, 66.7% Fraser and 61.2% Thomas). The chunks inserted into the REFERENT slots are wide-ranging; instantiations include simple nouns, *Mummy*, *people*, or *it*, article + noun constructions, such as *a train*, *the jigsaws* or *this one*, combinations of possessives + noun, for example, *your mouth* or *his tail*, as well as adjective + noun strings, *big books*, *yellow piece* or *funny noise*. The following slot-and-frame patterns were especially productive: *no [REF]*, *(a) big [REF]*, *what's (a) [REF]?*, *where's (a/the/my) [REF]?*, *I want [REF]*, *I see a [ATT] [REF]*, *(and/but) that's [REF]*, *I wanna [PRO] [REF]*, *let's go, [REF]* and *there's/there* *is (a/the) [REF]*. The dominance of REFERENT slots in recurring schemas as well as the abundance of semantically appropriate slot fillers puts forward that the children have already mastered the use of the noun phrase. Furthermore, the REFERENT slot appears to be at the source of much creativity at the age of 2;6. It is crucial to mention, however, that the REFERENT slot is only rarely filled with spatial terms, like shapes, *tower* and *tunnel* or locations, *this space* and *hospital*.

In addition, one can observe that the second largest group of constructions, comprising between 13.4 and 17.9 percent, possesses a PROCESS slot. In the schema *no [PRO]*, for instance, the slot is held by *fall down* or *fit in*, whereas the more abstract pattern *can you [PRO] [REF]?* requires the substitution of both *open* and *the door* to match the target utterance. Most slot fillers of the PROCESS category are infinitives (*let's [PRO]* > *let's go* and *you hafta [PRO] [LOC]* > *you hafta sleep there*) but the datasets also feature some present participle (*he's [PRO]* > *he's flying* and *[PRO] like that* > *doing like that*) or verb particles (shall I [PRO] [REF] [PRO]? > shall I pick this up?).

The percentages in Figure 13 moreover highlight that the children describe characteristics of entities and events; this is possible on the basis of ATTRIBUTE slots. For all three subjects, constructions with this slot account for less than 1 in 10 item-based schemas (9.5% Eleanor, 4.4% Fraser and 7.5% Thomas). Frequently, attributes are used in combination with referents, for example in *all the big books* (*all the [ATT] [REF]*) or *that's a funny noise* (*that's a [ATT] [REF]*). Other slot-and-frame patterns featuring an ATTRIBUTE slot are *that's [ATT]*, *[REF] all [ATT], too [ATT]*, *I see a [ATT] [REF]*, *[ATT] one* and *are you just [ATT]?*. Permissible slot fillers primarily relate to sizes, like *big*, *small*, *tall* and *little* (compare 'dimensions' in 5.1.), or colours, including *yellow*, *red* and *black*.

A peculiarity in the Eleanor-corpus is the use of a slot type designating time (3x), which has not yet been discussed in previous studies. The productions *come back another time*, *let's lie down for a bit* and *I'm not a big girl yet* follow the structure of the frames *come back [TIME]*, *let's [PRO] [TIME]* and *I'm not a [REF] [TIME]*, respectively. The material inserted spans a location in time, *another time*, static duration, *for a bit*, and adverb expressing the non-occurrence of a state, *yet*; in each case, the TIME slot occupies the construction-final position. Fraser and Thomas have presumably not abstracted across prior utterances to create such a slot, indicating that it might emerge only at a higher level of linguistic development.

From the beginning, this thesis has sought to investigate the intersection of construction learning and spatial language learning. LOCATION and DIRECTION slots in partially schematic patterns permit insights into both processes. The target utterances *I'm on it* in Table 21 and *he dropped it on the ground* in Table 22 demonstrate children's creative use of slot fillers pertaining to places at the age of 2;6:

Table 21. Tracing back *I'm on it* (Eleanor 020600a, 719)

Component units	Operations	Strategy	Attested frames or fixed strings	Attested slot fillers
I'm [LOC] on it	1	SUBSTITUTE	CHI: I'm upstairs. CHI: I'm in my car. CHI: I'm in the toilet. CHI: I'm outside [?]. CHI: I'm here.	CHI: don't go on it! CHI: it's got lions [?] on it. FAT: put paper on it and then you can draw.
Successful traceback: *I'm on it*				

Table 22. Tracing back *he dropped it on the ground* (Fraser 020601a, 1219)

Component units	Operations	Strategy	Attested frames or fixed strings	Attested slot fillers
[REF] dropped it he [LOC] on the ground	2	SUBSTITUTE SUBSTITUTE	CHI: I dropped it on the stairs. MOT: you dropped it on the stairs, didn't you?	CHI: he climbs up. MOT: it doesn't (be)cause he wriggles on the ground, doesn't he?
Successful traceback: *he dropped it on the ground*				

Based on prior representations of the schema *I'm [LOC]* found in the dataset, e.g. *I'm outside* or *I'm in my car*, Eleanor creates the novel production *I'm on it* by substituting *on it* into the LOCATION slot. Said slot is otherwise taken by the like lexemes *upstairs* and *here* or the prepositional phrase *in the toilet*, but never by the chunk *on it*. Yet, the slot filler can be traced back to other strings from which the child has ostensibly learned to use it in the *I'm [LOC]* schema as well.

The more abstract pattern *[REF] dropped it [LOC]* is realised as *he dropped it on the ground* and therefore lexically a close match with the precedents *I dropped it on the stairs* and *you dropped it on the stairs, didn't you?*. When Fraser applies the SUBSTITUTE operation to insert *he* into the REFERENT slot and *on the ground* into the LOCATION slot, the string can successfully be built from the main corpus. Accordingly, the utterance can be traced back because both the frame and the slot fillers are attested in earlier recordings.

While one tenth of Eleanor's slots in schematic constructions are of the LOCATION type, Fraser's and Thomas's use thereof is modestly higher at 13.3 and 16.4 percent. Apart from this, the data shows that material of different complexities fills the LOCATION slot, such as simpler *there* or *in here* and more advanced units like *in my bag* or *on that one*. It can be argued that productions featuring locations increasingly become less repetitive as children abstract across utterances and connect them in complex networks. Figure 14 is an attempt to visualise what the vertical and horizontal relationships between constructions might look like using the example of LOCATION slots:

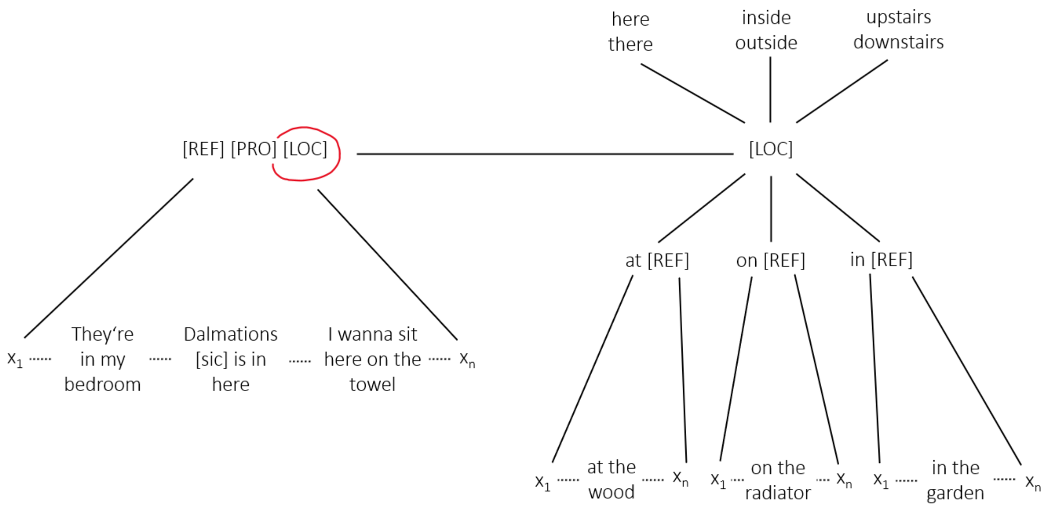


Figure 14. Network of lexically specific schemas and fully abstract patterns

I suggest that as children come across prepositional phrases in their input they recognise that the grammatical particle denoting a relation, *at*, *on* or *in*, is frequently followed by a referent, such as *the wood*, *the radiator* and *the garden*. Having already lexicalised the REFERENT slot, the children learn to create novel construction analogous to prior exemplars, using *at [REF]*, *on [REF]* and *in [REF]* as templates. All three schemas describe a location and thus contribute to the formation of a more abstract LOCATION slot, which is, however, initially lexically specific in itself. Yet, information about locations can also be conveyed through spatial adverbs, including *here*, *there*, *inside*, *outside*, *upstairs* and *downstairs*. With growing linguistic experience, children will categorise these linguistic structures as material permissible in the LOCATION slot as well. Finally, [LOC] can feature in constructions of different degrees of specificity. The utterances *they're in my bedroom*, *dalmations [sic] is in here* and *I wanna sit here on the towel*, all genuine language examples from the corpora, follow the high-level pattern *[REF] [PRO] [LOC]*, which the child will successfully master as soon as they have fully abstracted across instantiations without reliance on fixed lexical elements.

Lastly, I identified some utterances, albeit very few (3x Eleanor, 1x Fraser, 1x Thomas), exhibiting DIRECTION slots. The utterance *she's going to bed now* can be reconstructed through (a) superimposing *she* and (b) superimposing *to bed* into the open slots of the frame *[REF]'s going [DIR]* as well as (c) adding the temporal adverb *now* at the end. The traceback process is summarised in this table:

Table 23. Tracing back *&-um she's going to bed now* (Eleanor 020600a, 245)

Component units	Operations	Strategy	Attested frames or fixed strings	Attested slot fillers
[REF]'s going [DIR] she to bed + now	3	SUPER-IMPOSE SUPER-IMPOSE ADD	MOT: I think she's going outside. CHI: they [*] going to work. CHI: he's going to bed. MOT: are we going to the Trafford_Centre?	MOT: shall we go to bed? CHI: went to bed. MOT: she's going to xxx. CHI: she won't. CHI: it's gone on the tree now.
Successful traceback: *&-um she's going to bed now*				

The fact that LOCATION slots outbalance DIRECTION slots seems to hint at a tendency to convey information about the static positions of entities in space as opposed to perceived motion at the age under investigation. In all, this analysis has stressed the importance of variable slots in language acquisition.

5.3. Failed Derivations

In the previous section I have explained how the three children at the age of 2;6 use a slot-and-frame architecture to form novel spatial language constructions on the basis of prior ones. They can be built from the previous discourse by substituting, superimposing or adding schemas and by repeating fixed strings. The most obvious finding to emerge from the analysis is that between 70 (Eleanor) and 80 percent (Thomas) of the target utterances are reconstructable with the traceback approach. This, however, also highlights that about a quarter of them cannot fully be accounted for. An initial objective of this project (see RQ3) was therefore to identify reasons for failed derivations. In line with earlier research, this study, too, finds that derivations failed if either a word (lexical fail) or component unit (syntactic fail) that would have been required as a precedent was not represented in the main corpus. Figure 15 juxtaposes lexical and syntactic fails detected in the three datasets:

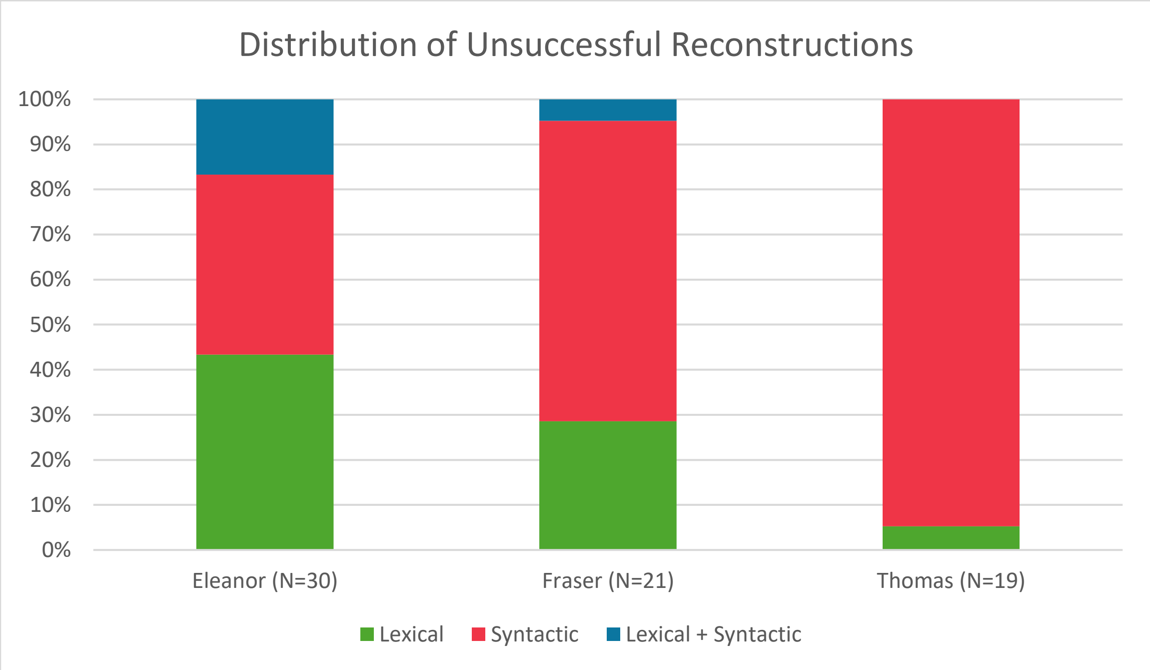


Figure 15. Relation of fails due to lexical versus syntactic reasons

According to this chart, about 43.3 percent of the failed derivations in the Eleanor-corpus, 28.6 percent in the Fraser-corpus and 5.3 percent in the Thomas-corpus can be attributed to the absence of individual words. Moreover, syntactic fails account for the majority of unsuccessful tracebacks, ranging from 4 (Eleanor) to 9 in every 10 (Thomas). A small number of utterances (5 in Eleanor 020600a, 1 in Fraser 020601a) cannot be reconstructed because of both lexical and syntactic reasons. Comparing the children at different levels of development, it appears that the proportion of lexical fails increases with higher MLU, whereas syntactic fails occur less frequently in more advanced learners.

In Tables 24 and 25 I present just three of the 20 constructions that could not be derived because of missing lexical elements:

Table 24. Tracing back *where's the fish fry?* and *where's Polly?* (Eleanor 020600a, 215 & 1072)

Component units	Attested frames or fixed strings
where's the [REF]? fish fry	INV: where's the ladybird? CHI: where's the ticket?
where's [REF]? Polly	CHI: where's Thomas? CHI: where's Daddy?
Failed traceback: missing lexical elements *fish fry* and *Polly*	

Table 25. Tracing back *that's not a hippopotamus* (Fraser 020601a, 1809)

Component units	Attested frames or fixed strings
that's not a [REF] hippopotamus	CHI: that's not a party. INV: that's not a doggy, is it MOT: that's not a basket.
Failed traceback: missing lexical element *hippopotamus*	

Where's the [REF], *where's [REF]* as well as *that's not a [REF]* are all highly productive schemas slots in Eleanor and Fraser's main corpora. As the examples above indicate, the open slots are filled with a variety of linguistic items, yet the referents *fish fry*, *Polly* and *hippopotamus* do not occur at all within the entire main corpora. That is not to say that the learners cannot possibly know these constructions; on the contrary, if a child produces a word, it must have picked it up somewhere before. *Hippo*, for instance, is represented 23 times in the Fraser-corpus (produced 12x by the mother, 8x by the child and 3x by the investigator). It is likely that the mother-child dyad just started using the more complex form *hippopotamus* around the time of the last recording, which means that the interval sampling technique did not capture it. Furthermore, I assume that Eleanor or her caregivers have also previously produced *fish fry* and *Polly,* but my limited dataset cannot account for any of those occurrences. Below I summarise all words that did not feature in the main corpora and thus caused lexical fails:

Table 26. Words found in the test corpora but not the main corpora

Eleanor	Fraser	Thomas
daddy_long_legs, fold, daughter, fish fry, Polly (5x), washing machine, types, dalmations [*sic*] (2x), dinner mat, hammock, upstanding, book work, pushchair	sliding (2x), spinning, Jack_in_the_box, hippopotamus (2x), rocket	platform

Even the strings that fail to be reconstructed due to lexical reasons can be read as evidence in favour of children's item-based language learning. They are arguably also created on the basis of attested partially productive patterns and, hence, despite the unavailability of certain words, closely connected to previous utterances.

In pursuing syntactic fails, I, like Lieven et al. (2009) and Koch (2019), further analyse different types of syntactically unsuccessful derivations. The first syntactic reason some target utterances could not be built from the main corpora was the use of incorrect slot fillers. The following example illustrates that Fraser inserts a word of the wrong semantic category into the established schema *I wanna [PRO] like [REF]*:

Table 27. Tracing back *I wanna botbot like that cow* (Fraser 020601a, 623)

Component units	Attested frames or fixed strings
I wanna [PRO] like [REF] botbot that cow	CHI: I wanna have a bath like Maisy. CHI: I wanna see it. CHI: that's a cow.
Failed traceback: incorrect slot filler *botbot [=baby bottle]* only used as [REF] and not [PRO]	

While all component units required for the utterance *I wanna botbot like that cow* exist in the dataset, it was nonetheless flagged as a syntactic fail on the grounds of the incompatibility of *botbot* and the PROCESS slot. *Botbot*, Fraser's babytalk for bottle, is only used as a referent and never extended to mean *to drink* (verb) as well. The lexically specific pattern at hand, however, does not allow substituting referents into the PROCESS slot since it is semantically restricted. *Botbot* would thus be eligible to fill the slot in *I want a [REF]*, for instance.

The production *it is there*, albeit perfectly correct, is also interpreted as a syntactic fail. On the one hand, the open slot following *it is* never takes a location in Eleanor's main corpus but only referents, like *a snake*, and attributes, such as *sharp*, as can be seen in the next table:

Table 28. Tracing back *it is there* (Eleanor 020600a, 141)

Component units	Attested frames or fixed strings
it is [LOC] there	CHI: it is a snake. CHI: it is sharp. CHI: there it is.
Failed traceback: incorrect slot filler *it is* only followed by [REF] or [ATT], never [LOC] OR alternative construction *there it is*	

Arguably, there is no evidence that the child has abstracted the schema *it is [LOC]* from previous utterances which rather exhibit the structures *it is [ATT]* and *it is [REF]*. The traceback method can therefore not derive the construction because *there* does not fulfil the semantic requirements of the slot. I cannot confirm that the child knows to invert the structure based on prior productions. On the other hand, an alternative account suggests that no match for the utterance can be created as there is no type variation to form the abstract pattern *it is [LOC]* with a location in the final position. Yet, a competing string, *there it is*, features in the Eleanor-corpus, suggesting that the LOCATION slot should be frame-initial.

Another explanation for syntactic fails pertains to strings which include extra linguistic elements otherwise not integrated in the schema or miss constituents that are part of it. For the sentence *I hurt on those track* to map onto the attested frame *I hurt [REF] on [REF]*, a reflexive pronoun or body are required in the first REFERENT slot.

Table 29. Tracing back *I hurt on those track* (Fraser 020601a, 141)

Component units	Attested frames or fixed strings
I hurt on [REF] those track	CHI: I hurt my foot on that. CHI: I hurt myself on the books. MOT: and that's Henry and he's waiting for them to get the cows off so_that he can go along the track.
Failed traceback: reflexive pronoun or body part missing *I hurt [REF] on [REF]*	

Owing to the restrictions of the ADD operation, however, new material cannot be inserted into the existing schema. ADD only allows the linear juxtaposition of constituents at the beginning or the end of a construction. Although this is clearly the case in the following example, where *in the bin* is attached at the end of the fixed string *look up*, it also counts as a syntactic fail according to the rules of traceback.

Table 30. Tracing back *look up in the bin* (Eleanor 020600a, 783)

Component units	Attested frames or fixed strings
look up + [LOC] + in the bin	FAT: look up. FAT: Dada's gonna [: going to] throw it in the bin now.
Failed traceback: fixed string *look up* cannot be combined with additional [LOC]	

The ADD operation is restricted to vocatives, conjunctions and adverbs; in the bin does not belong to either of these categories. Table 30 above shows how the utterance in question violates the restrictions set forth. I will address the plausibility of this in the discussion section.

The production *get in the rabbit* cannot be derived successfully since the word order is arranged in an unconventional manner ("Herausstellungsstruktur"; see Koch, 2019, pp. 251-252):

Table 31. Tracing back get in the rabbit (Eleanor 020600a, 783)

Component units	Attested frames or fixed strings
get in [?] [REF] the rabbit	MOT: the rabbit eats carrots. CHI: just get in your hair.
Failed traceback: word order mixed up: *the rabbit [PRO]* OR incorrect slot filler: *get in [LOC]*	

When Eleanor utters *get in*, she specifies that it is *the rabbit* who gets in right after. *The rabbit* is dislocated to the right in this example even though the referent usually precedes the PROCESS slot. The utterance hence neither matches the schema *the rabbit [PRO]* nor *get in [LOC]*. It becomes apparent that the traceback approach reaches its limits if the information structure of a sentence is rearranged.

While all the types of syntactic fails mentioned thus far are related to attested constructions on a lexical level, supporting the usage-based assumption of low-level schemas in early multiword speech, the last fail I identified is the product of wholly abstract patterns. The table below presents an example sentence for which no lexically specific schema was found:

Table 32. Tracing back *yellow truck come by* (Thomas 020602, 783)

Component units	Attested frames or fixed strings
[ATT] [REF] come by yellow truck	CHI: yellow envelope. CHI: a big truck comes now. INV: there's a train coming by. [>] CHI: big bears eat. CHI: big car smash.
Failed traceback: no fixed or schematic but only fully abstract pattern *[ATT] [REF] [PRO]*	

One expects to discover in the Thomas-corpus utterances that follow the pattern *[ATT] [REF] come by*, which would account for the target utterance *yellow truck come by*. While all component units to build the production, *yellow [REF]*, *[ATT] truck* and *coming by*, are represented, they do not occur together in partially schematic frames. Knowing that utterances like *big bears eat* and *big car smash* exist in the corpus, the child might abstract the pattern *[ATT] [REF] [PRO]* and substitute *yellow*, *truck* and *come by* into the respective slots. Yet, this abstract frame does not have a fixed lexical element, which is imperative for the traceback to be successful. The method can consequently only account for schematic but not fully abstract patterns.

The majority of Thomas's failed derivations, furthermore, did not fall into the categories of syntactic fails discussed above. For utterances that could not be built from the main corpus, it was largely impossible to discern what the child was trying to convey. I cannot access the meaning of constructions like *back fall down* or *no hands down* (Thomas 020602, 13 & 1277), which made extracting patterns and creating matches practically unfeasible. In sum, this section has emphasised the range of reasons, both lexical and syntactic, why some derivations fail, including the unavailability of certain words, the use of incorrect slot fillers, the presence and absence of additional constituents or the violation of restrictions of the ADD operation. The traceback method seemingly cannot account for fully abstract patterns in early multiword speech but fails to deal with high degrees of ambiguity, too.

6. Discussion

This present investigation, informed by construction grammar and usage-based theories of language acquisition, delivers promising results regarding the item-based nature of spatial utterances for Eleanor, Fraser and Thomas aged 2;6. In preparation for the spatial language traceback study at the heart of this thesis, I have categorised spatial terms according to the University of Chicago spatial language coding system (Cannon et al., 2007 cited in Ferrara et. al, 2011). Similarly to Yang and Pan (2021), who investigated Chinese kindergartners aged 3 to 5 during block play, I detect a large number of spatial locations and deictic terms in Eleanor and Fraser's speech. What is curious, however, is that Thomas uses 81, as opposed to 32 and 31, spatial orientations and transformations. A note of caution is due here since Thomas lags behind in development and tends to repeat parental input word for word. In fact, he exhibits linguistic behaviour observed by Tomasello (1987) in his subject at 1;5, using locative prepositions holophrastically or omitting them altogether. Beyond that, there seems to be a preference for static spatial relations (stasis) over dynamic ones (kinesis), possibly because the underlying concepts are easier to grasp (Piaget & Inhelder, 1967; Johnston & Slobin, 1979). I have, moreover, established the early mastery as well as the relevance of spatial terms for children at the multiword stage (Wu et al., 2022). It remains to be seen, however, if children, similarly to the distribution and production rates of parts-of-speech investigated by Behrens (2006), approximate the adult system of spatial language as modelled by their caregivers; the input-output relationship of linguistic representations of space was beyond the scope of this thesis.

In reviewing the literature, no data was found on the association between spatial language acquisition and slot-and-frame patterns in early multiword speech. This thesis thus set out to apply the traceback method, developed "to account for all of the utterances produced in a short test corpus […] using lexically specific schemas and holophrases attributed to the child on the basis of previous sessions" (Ambridge & Lieven, 2011, p. 218), to constructions including references to space. Several earlier reports have concluded that 90 percent (Dąbrowska & Lieven, 2005), 85-90 percent (Lieven et al., 2009) or 60-89 percent of English-speaking children's constructions (Kol et al., 2014) can be accounted for by tracing back to attested child utterances and child-directed speech. Both Miorelli (2017) and Koch (2019) in examining one Italian child and four German-speaking subjects, respectively, confirm that close to all strings, between 82 and 95 percent to be exact, can be built from the main corpora. Although the target utterances reconstructed here have been restricted considerably to reflect only the children's repertoire of spatial constructions, my findings accord with previous observations. At 70.0% (Eleanor), 77.7% (Fraser) and 79.6% (Thomas), the proportions of successful derivations are largely consistent – in fact, only marginally lower – with those of the seminal studies.

Due to the limited dataset, determining how much of the children's talk is actually new, i.e. has never been produced in this form before, proves to be a challenge. It is possible, however, give some indication as to which productions are definitely not new because they already occur in the corpus. Between 16 and 25 constructions per child have been classified as exact matches, suggesting that a notable share of productions at 2;6 are indeed fixed strings. That is not to say that they cannot be the result of creative processes, too. Yet, in line with Diessel (2011, p. 834) I contend that "memory is cheap, computation is costly"; children rely on memory, not fully abstract rules, and would select the most efficient route to sentence formation (Dąbrowska & Lieven, 2005). It is fair to assume that repetitions require less brainpower, so, wherever possible, strings are acquired as preconstructed chunks. The frozen phrases detected in the three datasets are succinct and, most notably, linked to routinised contexts. Following Goldberg (2006), I postulate that they occur at high frequencies in the children's linguistic environments and thus become fully predictable. The present study, therefore, corroborates the significance of fixed formulae posited not only in traceback studies (Lieven et al., 2003; Dąbrowska & Lieven, 2005; Lieven et al., 2009) but also in other works on rote-learning and specificity in early speech (Pine & Lieven, 1993; Tomasello, 2003; Bannard & Matthews, 2008).

As hypothesised in section 4, however, the vast majority of constructions are organised around specific words and consequently the product of schemas with open slots. Dąbrowska and Lieven (2005), for instance, report that 55 and 66 percent of child utterances at 2;0 can be reconstructed from the corpus using a single operation. In analysing children who are six months older, the present study finds that a marginally smaller, yet substantial share of 38.6, 42.5 and 51.4 percent of target productions, depending on dataset, require only one modification to arrive at a match. Moreover, multi-operation tracebacks (two operations and 3 or more operations) account for 38.6 (Eleanor), 26.0 (Fraser) and 14.9 percent (Thomas) of successful derivations, compared to an average of 13.8 percent in Lieven et al. (2009) and 33.8 in Koch (2019). These findings may be taken to indicate that early multiword speech is closely related to prior constructions on a lexical level and revolves around schemas with open slots.

Yet, children are not born with mental representations of such schemas; they emerge with growing exposure to linguistic structures. According to Tomasello (2003; 2006), learners expand on the general-purpose cognitive skill of pattern-finding by recognising similarities and differences in their input. They abstract across instantiations and integrate new exemplars into already established patterns, which become entrenched upon frequent re-occurrence (Langacker, 2000; Goldberg, 2006). The prevalence of fixed formulae and schematic constructions in the three children's datasets lends credence to the learning processes recommended in the theory section and stresses that item-based schemas are prioritised over fully abstract, high-level patterns. At the same time, this casts serious doubt on the generative understanding of language acquisition; for decades, generativists have claimed that early structures are too creative to be learnable form exposure alone and must hence be innate (Chomsky, 1965; 1980; Pinker, 1994). In fact, however, with the help of the traceback method, It can be argued that children's language use at this age is, on the one hand, strongly influenced by parental input and, on the other hand, quite conservative, exhibiting only constrained creativity, i.e. where open slots permit variation.

One previous study (Lieven et al., 2009) has, additionally, described a connection between complexity of child utterances, as measured in MLU, and operations required to derive them from prior productions. The authors propose that exact repeats (zero-operation) decrease with higher MLU but the likelihood of multi-operation tracebacks increases with MLU (Lieven et al., 2009, p. 492). While Koch (2019) only corroborates this correlation for 3 of 4 children in his German traceback study, I observe this link in all three children. The Thomas-corpus (MLU 3.400) features most verbatim repetitions, yet no tracebacks requiring more than 2 operations. Eleanor and Fraser (MLU 6.243 and 5.480) in turn produce less frozen phrases and a larger number of frames with at least one slot, which Dąbrowska and Lieven (2005, p. 452) attribute to a bigger working memory allowing the children to more productively combine and reassemble linguistic units. My results arguably further support Lieven et al.'s (2009, p. 503) statement that "[t]he degree of repetitiveness […] found suggests that children are initially learning not only words but strings as 'big words' which subsequently start to be internally analyzed". The children investigated in my thesis are still at an early stage of their linguistic development. They have not yet mastered fully abstract patterns and crucially depend on multiword sequences that are first used holistically before they are analysed into smaller component parts.

I therefore argue that my findings also strengthen the Starting Big approach, stating that "[a]s they accumulate linguistic experience, children will segment these initially under-segmented chunks" (Arnon, 2021, p. 940). As mentioned in review section, usage-based linguists have noted the prevalence of multiword units in form of frequent frames in the input children receive (Cameron-Faulkner et al., 2003; Mintz, 2003); by relating target utterance, inter alia, to child-directed speech, I, too, strengthen the argument that young minds acquire grammatical knowledge from the constructions modelled by their caregivers. Based on the test corpora it becomes clear that the three children produce individual spatial terms (e.g. *there* or *fit*) and longer strings featuring references to space (e.g. *where's Daddy?* or *I wanna play that*). I have illustrated how they abstract across such usage events to create partially realised schemas with open slots (e.g. *where's [REF]?* or *I wanna [PRO] [REF]*). Evidently, this thesis adds to the existing research viewing multiword strings as integral building blocks and thereby challenges the primacy of words in language acquisition (Christiansen & Arnon, 2017; Theakston & Lieven, 2017; Arnon, 2021). It did not, however, expound the exact cognitive mechanisms underlying the acquisition of abstract knowledge, which is why further studies will need to be undertaken.

Research question 1 also asked which operations link such schematic constructions in the test corpus to the main corpus. Following Koch's (2019) proposal, constructions were built with the operations SUBSTITUTE, SUPERIMPOSE and ADD. The relative dominance of substitutions and superimpositions highlights the relevance of item-based schemas in construction learning; conversely, ADD accounted for only 3.7 (Eleanor) or 13.2 percent of the utterances. It is encouraging to compare these figures to Lieven et al. (2009, p. 494) and Koch (2019, p. 213) who determined that the ADD operation was only applied in 5 or 0.5 to 6 percent of successful reconstructions, respectively. I have already explained that the use of juxtapositions is highly restricted, which lead to some failed derivations. Langacker (2008, p. 221), for instance, criticises the assumption that schemas are governed by inherent restrictions:

> As ready-made resources available for exploitation, schemas function as attractors within [the space of structured possibilities], thus inhibiting the exploration of other areas. In this way, the positive characterization of conventional patterns can indicate implicitly (and quite effectively) that options outside their range are nonconventional and will be judged ill-formed.

He reasons that schemas are positive specifications and cannot possibly convey which expressions are ungrammatical and thus not permissible based on some implicit rules. It is, therefore, psychologically implausible for children to know which material to add without breaking said rules of language. In fact, children learn which linguistic structures are unacceptable through non-occurrence and not prohibitions. Langacker's discussion coincides with Goldberg's (2011) concept of statistical pre-emption and Lieven's (2010) elaborations on frequency effects. Future investigations might consider relaxing the restrictions of the ADD operation and incorporating these linguists' deliberations into an updated version of the traceback approach (see also Koch, 2019, p. 265). Yet, it is important to note here that the restrictions of traceback are intended as instructions for linguists to constrain the approach, and not as rules imposed on the child.

I have, moreover, shown that the traceback approach is by no means exhaustive. For various reasons, about one fifth to one third of the target utterances could not be derived from prior productions. For some constructions, no successful match was created because individual words were simply not attested in the corpus; others were classified as syntactic fails, due to the absence of suitable frames, for example. Dąbrowska and Lieven (2005, p. 435) put forward that failed derivations are likely "an artifact of sampling". The authors write that the main corpora, partial samples of children's linguistic environment, might not have captured the lexeme or lexically specific pattern in question. The data sample used in this study is "intensive" (Lieven & Goh, n.d.-a; Lieven & Goh, n.d.-b), covering the first week of the month over a 2-months recording period, but not "dense", like many child development researches call for (e.g. Lieven & Behrens, 2012). This limitation surely explains why not all child productions were accounted for. It possibly also explains the lower share of successful tracebacks for more advanced children, like Eleanor, as they exhibit more diverse vocabulary and slot-and-frame patterns that might not be featured in the recordings.

In evaluating the traceback approach, Koch (2019) questions whether the seemingly random selection of (size of) test and main corpus allows for reliable results. Following systematic criteria (see section 4.1.), I selected between 16 and 20 recordings from the two months before the final session that serves as a test corpus; the Eleanor-corpus comprises 16:05:29 hours of parent-child interaction whereas the Fraser-corpus and Thomas-corpus are slightly bigger (20:54:12 and 18:18:04 hours). A child is awake and experiencing language for an estimated 10 hours a day; traditional acquisition studies can only account for only 1 percent, whereas even dense databases capture 10 percent of their linguistic environment (Lieven et al., 2003, p. 336; Tomasello & Stahl, 2004, p. 102; Rowland et al., 2007, p. 2). Over a time span of 2 months, the children analysed in this present project were presumably producing or perceiving speech for 600 hours each (10 hours/day for 60 days). A small corpus of 16 to 21 hours logically only encloses 2.7 to 3.5 percent of their language experience. I can moreover not be sure if the precedents I identified are actually the ones from which the target expression was derived. I am aware that these factors reduce the explanatory power of my study.

Yet, this shortcoming cannot be overcome by simply extending the dataset since speech samples can never be fully representative of children's linguistic experience. As linguists have pointed out, this is not a problem specific to traceback but one that concerns the entire field of language acquisition research (Koch, 2019). Constructions which did not match antecedents in the corpora can ostensibly be interpreted as evidence that children use particularly creative structures surpassing the limits of the traceback method. Assuming that all child productions must be, in theory, derivable, contradicts the view of a resourceful learner (Koch, 2019, p. 265). Koch (2019, p. 265) remarks that children might not strive to producing only strings parallel to what they already know to be in line with the grammar of their target language ("zielsprachige Grammatik").

This thesis was also designed to determine which slots in schematic constructions are relevant to utterances pertaining to space. I find that more than half of the slots in lexically specific schemas are of REFERENTS, which is roughly consistent with Koch's (2019) 47.2 to 60.6 percent but considerably lower than Lieven et al.'s (2009) 61 to 93 percent. Incidentally, nouns are also among the first-acquired words (Tardif et al., 2008), which gives children a longer practice period to learn their usage patterns. Eleanor, Fraser and Thomas display good command of the noun phrase as the linguistic material inserted in the REFERENT slot includes simple nouns and pronouns (*Mummy* or *she*) but also determiner + noun (*a train* or *this one*) or adjective + noun constructions (*big books* or *yellow piece*). Hence, it could conceivably be hypothesised that the children of this study successfully compare new exemplars of referents with known representatives of the category, thus establishing productive schemas (e.g. Langacker, 1987). PROCESS and ATTRIBUTE slots feature less frequently; processes often occur as infinitives and follow auxiliaries (*I can [PRO] it* or *don't [PRO] it*) or semi-auxiliaries (*I wanna [PRO]* or *let's [PRO]*) (Lieven et al., 2009, p. 496). Whereas Koch (2019, p. 241) maintains that attributes are largely used predicatively, therefore being organised around the verb *to be* (German 'sein'), I find that they are, for the most part, used attributively in my datasets and thus closely linked to nouns (*black leg* or *big baby*).

Arguably, the hypothesis that REFERENT, PROCESS and ATTRIBUTE slots are well-represented in the target utterances (see research question 2, hypothesis II) was also corroborated. This expectation, however, was rooted in the misperception that it would be the spatial terms, i.e. shapes, spatial orientations and transformations or dimensions, that take the unrealised slots. In fact, more often than not the spatial terms were part of the lexically specific frames and not inserted into the constructions' variable positions (*where's [REF] going?*, *hiding behind [REF]* , *a big [REF]* or *open the [REF]*). Still, this rather intriguing finding emphasises that spatial language learning, like early construction learning at large, is item-based in nature.

Like Weist et al. (2000) and Meints et al. (2002) I find that simple locative prepositions are produced by children already after their second birthday. Yet I can only partially verify the hypothesis that locations and directions feature prominently in the three datasets. As the test corpora, limited to linguistic references to space, are a concentrated pool of spatial terms, it is not surprising that the proportion of LOCATION slots, 10 percent for Eleanor, 13.3 for Fraser and 16.4 for Thomas, is substantially higher than that reported in other traceback studies. DIRECTION slots, on the other hand, occur equally infrequently across all investigations, including my own (Lieven et al., 2009; Koch, 2019). The present study raises the possibility that they are not as common as initially anticipated since spatial information is also encoded in chunks other than locations and directions. An alternative explanation might be that said structures are not as numerous in children's input, which makes recognising and generalising across the variable elements occupying LOCATION or DIRECTION slots in constructions challenging. Dąbrowska and Lieven (2005, p. 448), too, clarify that "these categories are [not] available to the child from the very beginning: [...], they emerge gradually from the child's linguistic experience and may initially be construction-specific". For convenience, the same labels REFERENT, PROCESS, ATTRIBUTE, LOCATION, DIRECTION and TIME have been used to discuss the slots in different constructions.

A small excursus in the analysis section was dedicated to the interconnectedness of schemas. Langacker's (2008) theoretical discussion of networks of schemas is backed by the genuine language data I presented. He proposes that language itself can be seen as "an enormous inventory of conventional linguistic units, all abstracted from usage events" (Langacker, 2008, p. 221). Schemas are always immanent in instantiations of these linguistic units, whether they are distinct or schematic in themselves. It is worth noting that they are not separate of each other but form structured assemblies through categorisation (Langacker, 2008, pp. 221-222). Using the example of the LOCATION slot, I argued that exemplars are connected and feed into more abstract patterns, like *[REF] [PRO] [LOC]*. There is abundant room for further progress in determining how constructions come together to build increasingly complex and abstract networks.

Taken together, this thesis yields interesting results concerning spatial language learning and schematisation in language acquisition. I have drawn attention to the implications of the study's evident weaknesses on my interpretation of early schematic constructions. In what follows, I again address the hypotheses formulated and recapitulate if they have been (partially) corroborated or rejected:

Table 33. Validity of the hypotheses proposed at the beginning of the study

Research Question 1		Research Question 2		Research Question 3	
Hypothesis I	✓	Hypothesis I	≈	Hypothesis I	✓
Hypothesis II	✓	Hypothesis II	≈	Hypothesis II	✓
Hypothesis III	✓	Hypothesis III	✓		
Hypothesis IV	✓				

✓ … corroborated, ≈ … partially corroborated, ✗… rejected

With respect to research question 1, I have arguably presented evidence in support of all four hypotheses. More than two thirds of the children's target utterances were successfully derived from what they previously heard or produced. In accordance with prior traceback studies, this investigation has highlighted that novel constructions are lexically related to prior strings, being either identical or only slightly modified through substitution, superimposition and juxtaposition. The complexity and abstractness of child productions is affected by MLU.

In hindsight, I have overestimated the role of LOCATION and DIRECTION slots, which are incommensurate with REFERENT slots. I have, however, shown that frames are created and activated when productions share features. Langacker (2000, p. 106), therefore, contends that "lower-level schemas, i.e. structures with greater specificity, have a built-in advantage in the competition with respect to higher-lever schemas". In this regard, spatial constructions follow the same process of abstraction and schematisation as constructions without linguistic representations of space.

Between 20 and 30 percent of constructions per child could not be built from the main corpora. The failed derivations are a product of the datasets and general methodological limitations; nonetheless, they "appear to be extensions of well-attested patterns rather than utterances requiring a radically different kind of knowledge" (Dąbrowska & Lieven, 2005, p. 464). I thus find both hypotheses associated with research question 3 to be supported.

7. Conclusion

The aim of this thesis was to determine how much of early multiword speech pertaining to space can be related to prior constructions by applying the traceback method. In section 5 I explained children's use of different operations, fixed strings and partially productive schemas with slots. My interest in spatial language acquisition originally grew from the debate on the precise learning mechanisms enabling children to become competent and productive speakers. After years of dominance by generative theories (Chomsky, 1980; Crain, 1991; Pinker, 1994), usage-based linguists have challenged assumptions of innate principles and a universal grammar in favour of an account of language acquisition which emphasises the significance of social interaction and domain-general cognitive skills in linguistic development (Tomasello, 2003; Hoff, 2006; Christiansen & Chater, 2008).

The traceback method (e.g. Dąbrowska & Lieven, 2005; Lieven et al., 2009; Koch, 2019) builds on this emergentist understanding of language acquisition to show that children's early multiword utterances can be built from prior productions, both by the child itself or the caregivers, and is thus more stereotypical than previously suggested. In the past, however, structures featuring references to space remained unexplored in traceback investigations; this present study was therefore undertaken to cast light on the formation and schematisation process of spatial language constructions. In order to assess the degree of lexical specificity of the target utterances under analysis, I compiled three small corpora on the basis of existing ones (MPI-EVA-Manchester Corpus [Lieven & Goh, n.d.-a]; Thomas Corpus [Lieven & Goh, n.d.-b]) donated to the computerised child language database CHILDES (MacWhinney & Snow, 1985).

Following the traceback procedure, this research has linked productions featuring spatial terms, as classified by Ferrara et al. (2011) and Yang and Pan (2021), to prior utterances attested in preceding recording sessions. I have successfully traced back 7 to 8 of 10 target strings, of which 22.9 (Eleanor), 31.5 (Fraser) and 33.8 percent (Thomas) were verbatim matches and 27.1 (Eleanor), 24.7 (Fraser) and 37.8 percent (Thomas) required only one modification. The investigation has also found that generally MLU made a considerable difference to the use of zero- and multi-operations. REFERENT, PROCESS, ATTRIBUTE, LOCATION, DIRECTION and TIME emerged as prominent slots in schemas; locations and directions, albeit more frequent than in other traceback studies, represent only a small share of slots in all three children. Yet all things considered, my findings clearly substantiate that children lean on "utterance templates" (Ambridge & Lieven, 2011, p. 124) to facilitate assembling new strings. Therefore, I claim that spatial constructions are item-based at 2;6 and become more abstract with increased linguistic experience.

Another aim of this thesis was to critically reflect upon the potential and shortcomings of traceback on a meta-methodological level. Results confirm that, on average, a quarter of child utterances cannot be matched to precedents. I have offered a number of diagnoses, including lexical and syntactic reasons, as to why some reconstructions were unsuccessful. First, the datasets constitute only partial excerpts of Eleanor, Fraser and Thomas's linguistic environments and, subsequently, might not have captured the children's use of otherwise productive words or schemas. The traceback approach was evidently not designed to account for fully abstract patterns or utterances exceeding the linguistic creativity attested in the main corpora either. Moreover, Thomas (MLU 3.400) might, to some extent, not have been advanced enough for a fruitful application of the traceback approach. While traceback was intended for multiword speech, he still seems to be transitioning from the telegraphic stage. Corrigan (2011, p. 280) proposes selecting subjects based on their mean length of utterance, not their age; this would allow linguists to better examine similarities in development instead of individual, potentially MLU-related, differences. Having demonstrated that MLU is a more reliable predictor of complexity and derivability of strings, I argue that this selection criterion might also be helpful in countering some of the aforementioned weaknesses of the traceback method.

Besides the limited size of my corpora and the shortcomings inherent in the method applied, I also question the cognitive plausibility of some aspects of the traceback approach. There is uncertainty in the research community whether the operations SUBSTITUTE, SUPERIMPOSE and ADD correspond to the actual mechanisms for creating novel constructions. Built-in restrictions, especially those of the ADD operation, contradict the view that learners acquire linguistic units not through prohibitions, but on the basis of convention as modelled in their input (Langacker, 2008; Goldberg, 2011). Additionally, just because the children have produced or heard a structure a few times before, it is not guaranteed that they established a mental representation thereof. If the discussion is to be moved forward, a better understanding of the neurological bases of construction learning should be developed.

Nonetheless, this current study can be regarded a valuable and necessary contribution to the literature associated with construction learning at the early multiword stage as well as the body of work focusing on children's spatial language repertoires, which have not yet been elucidated in synergy. My aim was to bring early spatial terms to the fore, which impact literacy and mathematical skills (Miller et al., 2016). To promote children's development, caregivers and educators ought to actively model and encourage the use of spatial language. A natural progression of this work is to analyse language of space and its lexical specificity in dense samples (e.g. Lieven & Behrens, 2012) covering more children over an extended time. I have only investigated three subjects, who incidentally overlap with the children participating in other traceback studies. I might inadvertently have perpetuated their idiosyncrasies; to underpin my findings and enhance their generalisability, a project similar to this one should be carried out with a different group of children. More information on the internal constituency of slot fillers denoting location and direction is needed to further explore schematic constructions of space and their slots. Future studies must also provide more definitive evidence on the process of abstraction and formation of the "construct-i-con" (Hilpert, 2014, p. 13) in young language learners.

8. References

Ambridge, B., & Lieven, E. (2011). *Child language acquisition: Contrasting theoretical approaches*. Cambridge University Press. https://doi.org/10.1017/CBO9780511975073

Ambridge, B., Pine, J. M., & Lieven, E. (2014). Child language acquisition: Why universal grammar doesn't help. *Language, 90*(3), e53–e90. https://doi.org/10.1353/lan.2014.0051

Arnon, I. (2021). The Starting Big approach to language learning. *Journal of Child Language*, *48*(5), 937–958. https://doi.org/10.1017/S0305000921000386

Arnon, I., & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language, 62*(1), 67–82. https://doi.org/10.1016/j.jml.2009.09.005

Bannard, C., & Lieven, E. (2009). Repetition and reuse in child language learning. In R. Corrigan, E. A. Moravcsik, H. Ouali, & K. Wheatley (Eds.), *Formulaic language: Volume II: Acquisition, loss, psychological reality, functional explanations* (pp. 297–321). John Benjamins Publishing.

Bannard, C., & Matthews, D. (2008). Stored word sequences in language learning: The effect of familiarity of children's repetition of four-word combinations. *Psychological Science, 19*(3), 241–248. https://doi.org/10.1111/j.1467-9280.2008.02075.x

Behrens, H. (2006). The input-output relationship in first language acquisition. *Language and Cognitive Processes, 21*(1–3), 2–24. https://doi.org/10.1080/01690960400001721

Behrens, H. (2008). Corpora in language acquisition research: History, methods, perspectives. In H. Behrens (Ed.), *Corpora in language acquisition research: History, methods, perspectives* (pp. xi–xxx). John Benjamins Publishing. https://doi.org/10.1075/tilar.6.03beh

Behrens, H. (2009). Konstruktionen im Spracherwerb. *Zeitschrift für Germanistische Linguistik, 37*(3), 427–444. https://doi.org/10.1515/ZGL.2009.030

Behrens, H. (2011). Erstspracherwerb. In L. Hoffman & U. Quasthoff (Eds.), *Die Matrix der menschlichen Entwicklung* (pp. 252–273). De Gruyter. https://doi.org/10.1515/9783110259766.252

Bennett, D. C. (1975). *Spatial and temporal uses of English prepositions: An essay in stratificational semantics*. Longman.

Boroditsky, L. (2000). Metaphoric structuring: Understanding time through spatial metaphors. *Cognition, 75*(1), 1–28. https://doi.org/10.1016/S0010-0277(99)00073-6

Boroditsky, L. (2012). How the languages we speak shape the ways we think: The FAQs. In M. Spivey, K. McRae, & M. Joanisse (Eds.), *The Cambridge handbook of psycholinguistics* (pp. 615–632). Cambridge University Press. https://doi.org/10.1017/CBO9781139029377.032

Bowerman, M. (1974). Learning the structure of causative verbs: A study in the relationship of cognitive, semantic, and syntactic development. *Papers and Reports on Child Language Development, 8*, 142–178.

Bowerman, M. (1975). Cross-linguistic similarities at two stages of syntactic development. In E. H. Lenneberg & E. Lenneberg (Eds.), *Foundations of language development: A multidisciplinary approach* (pp. 267–282). Academic Press. https://doi.org/10.1016/b978-0-12-443701-2.50026-3

Bowerman, M. (1996). The origins of children's spatial semantic categories: Cognitive versus linguistic determinants. In J. J. Gumperz & S. C. Levinson (Eds.), *Rethinking linguistic relativity* (pp. 145–176). Cambridge University Press.

Bowerman, M., & Choi, S. (2003). Space under construction: Language-specific spatial categorization in first language acquisition. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and thought* (pp. 387–427). MIT Press. https://doi.org/10.7551/mitpress/4117.003.0021

Brown, P. (2015). Space, linguistic expression of. In J. D. Wright (Ed.), *International encyclopedia of the social and behavioral sciences: Volume 23* (2nd ed., pp. 89–93). Elsevier. https://doi.org/10.1016/B978-0-08-097086-8.57017-2

Brown, R. (1973). *A first language: The early stages*. Harvard University Press.

Burgess, N., Jeffery, K., & O'Keefe, J. (1999). *The hippocampal and parietal foundations of spatial cognition*. Oxford University Press.

Bybee, J. (2006). From usage to grammar: The mind's response to repetition. *Language, 82*(4), 711–733. https://doi.org/10.1353/lan.2006.0186

Bybee, J. (2010). *Language, usage, and cognition*. Cambridge University Press. https://doi.org/10.1017/CBO9780511750526

Cameron-Faulkner, T., Lieven, E., & Tomasello, M. (2003). A construction based analysis of child directed speech. *Cognitive Science, 27*(6), 843–873. https://doi.org/10.1207/s15516709cog2706_2

Campbell, R. N. (2006). Pre-scientific studies of language development. In K. Brown (Ed.), *Encyclopedia of language and linguistics* (2nd ed., pp. 391–394). Elsevier.

Carlson, L., & van der Zee, E. (2005). Functional features in language and space. In L. Carlson & E. van der Zee (Eds.), *Functional features in language and space: Insights from perception, categorization, and development* (pp. 1–8). Oxford University Press.

Casasola, M., & Cohen, L. B. (2002). Infant categorization of containment, support and tight-fit spatial relationships. *Developmental Science*, *5*(2), 247–264. https://doi.org/10.1111/1467-7687.00226

Choi, S., & Bowerman, M. (1991). Learning to express motion events in English and Korean: the influence of language-specific lexicalization patterns. *Cognition*, *41*(1–3), 83–121. https://doi.org/10.1016/0010-0277(91)90033-Z

Chomsky, N. (1957). *Syntactic structures*. Mouton.

Chomsky, N. (1965). *Aspects of the theory of syntax*. MIT Press.

Chomsky, N. (1980). Rules and representations. *The Behavioral and Brain Sciences*, *3*(1), 1–15. https://doi.org/10.1017/S0140525X00001515

Christiansen, M. H., & Arnon, I. (2017). More than words: The role of multiword sequences in language learning and use. *Topics in Cognitive Science, 9*(3), 542–551. https://doi.org/10.1111/tops.12274

Christiansen, M. H., & Chater, N. (2008). Language as shaped by the brain. *The Behavioral and Brain Sciences*, *31*(5), 489–509. https://doi.org/10.1017/S0140525X08004998

Clark, E. V. (1973). What's in a word? On the child's acquisition of semantics in his first language. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language* (pp. 65–110). Academic Press.

Clark, E. V. (2016). *First language acquisition* (3rd ed.). Cambridge University Press.

Clark, E. V. (2019). Research on first language acquisition: A brief history. In J. S. Horst & J. v. K. Torkildsen (Eds.), *International handbook of language acquisition* (pp. 3–19). Routledge.

Corrigan, R. (2011). Using the CHILDES database. In E. Hoff (Ed.), *Research methods in child language* (pp. 271–284). Wiley. https://doi.org/10.1002/9781444344035.ch18

Crain, S. (1991). Language acquisition in the absence of experience. *The Behavioral and Brain Sciences*, *14*(4), 597–612. https://doi.org/10.1017/S0140525X00071491

Croft, W. (2001). *Radical construction grammar: Syntactic theory in typological perspective*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198299554.001.0001
Croft, W., & Cruse, A. D. (2004). *Cognitive linguistics*. Cambridge University Press. https://doi.org/10.1017/CBO9780511803864
Dąbrowska, E. (2000). From formula to schema: The acquisition of English questions. *Cognitive Linguistics, 11*(1–2), 83–102. https://doi.org/10.1515/cogl.2001.013
Dąbrowska, E., & Lieven, E. (2005). Towards a lexically specific grammar of children's question constructions. *Cognitive Linguistics*, *16*(3), 437–474. https://doi.org/10.1515/cogl.2005.16.3.437
Darwin, C. (1877). A biographical sketch of an infant. *Mind, 2*(7), 285–294. https://doi.org/10.1093/mind/os-2.7.285
De Boysson-Bardies, B., & Vihman, M. M. (1991). Adaptation to language: Evidence from babbling and first words in four languages. *Language*, *67*(2), 297–319. https://doi.org/10.1353/lan.1991.0045
Demuth, K. (2008). Exploiting corpora for language acquisition research. In H. Behrens (Ed.), *Corpora in language acquisition research: History, methods, perspectives* (pp. 199–205). John Benjamins Publishing. https://doi.org/10.1075/tilar.6.10dem
De Saussure, F. (1916). *Cours de linguistique générale*. Payot.
Diessel, H. (2009). Corpus linguistics and first language acquisition. In A. Lüdeling & M. Kytö (Eds.), *Corpus linguistics: Volume 2* (pp. 1197–1212). De Gruyter. https://doi.org/10.1515/9783110213881.2.1197
Diessel, H. (2011). Language, usage and cognition by Joan Bybee: Review. *Language, 87*(4), 830–844.
Diessel, H. (2013). Construction grammar and first language acquisition. In T. Hoffmann & G. Trousdale (Eds.), *The Oxford handbook of construction grammar* (pp. 347–364). Oxford University Press.
Elman, J. L., Bates, E., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness: A connectionist perspective on development*. MIT Press. https://doi.org/10.7551/mitpress/5929.001.0001
Ferrara, K., Hirsh-Pasek, K., Newcombe, N. S., Golinkoff, R. M., & Lam, W. S. (2011). Block talk: Spatial language during block play. *Mind, Brain and Education, 5*(3), 143–151. https://doi.org/10.1111/j.1751-228X.2011.01122.x
Fillmore, C., Kay, P., & O'Connor, C. (1988). Regularity and idiomaticity in grammatical constructions: The case of let-alone. *Language, 64*(3), 501–538. https://doi.org/10.2307/414531
Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2021). *Variability and consistency in early language learning: The Wordbank Project.* MIT Press. https://doi.org/10.7551/mitpress/11577.001.0001
Gaskins, D., Backus, A., & Quick, A. E. (2019). Slot-and-frame schemas in the language of a Polish- and English-speaking child: The impact of usage patterns on the switch placement. *Languages, 4*(8), 1–26. https://doi.org/10.3390/languages4010008
Goldberg, A. E. (1995). *Construction: A construction grammar approach to argument structure*. The University of Chicago Press.
Goldberg, A. E. (2003). Constructions: A new theoretical approach to language. *Trends in Cognitive Sciences*, *7*(5), 219–224. https://doi.org/10.1016/S1364-6613(03)00080-9
Goldberg, A. E. (2006). *Constructions at work: The nature of generalization in language*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199268511.001.0001

Goldberg, A. E. (2009). Constructions work. *Cognitive Linguistics, 20*(1), 201–224. https://doi.org/10.1515/COGL.2009.013
Goldberg, A. E. (2011). Corpus evidence of the viability of statistical preemption. *Cognitive Linguistics*, *22*(1), 131–153. https://doi.org/10.1515/cogl.2011.006
Goldberg, A. E. (2013). Constructionist approaches. In T. Hoffmann & G. Trousdale (Eds.), *The Oxford handbook of construction grammar* (pp. 15–31). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195396683.013.0002
Goldberg, A. E. (2019). *Explain me this: Creativity, competition, and the partial productivity of constructions*. Princeton University Press. https://doi.org/10.1515/9780691183954
Goldberg, A. E., Casenhiser, D. M., & Sethuraman, N. (2004). Learning argument structure generalizations. *Cognitive Linguistics, 15*(3), 289–316. https://doi.org/10.1515/cogl.2004.011
Grimm, H. (1975). On the child's acquisition of semantic structure underlying the wordfield of prepositions. *Language and Speech*, *18*(2), 97–119. https://doi.org/10.1177/002383097501800201
Hartmann, S., Koch, N., & Quick, A. E. (2021). The traceback method in child language acquisition research: Identifying patterns in early speech. *Language and Cognition*, *13*(2), 227–253. https://doi.org/10.1017/langcog.2021.1
Hilpert, M. (2014). *Construction grammar and its application to English*. Edinburgh University Press.
Hoff, E. (2006). How social contexts support and shape language development. *Developmental Review, 26*(1), 55–88. https://doi.org/10.1016/j.dr.2005.11.002
Hoffmann, T., & Trousdale, G. (2013). Construction grammar: Introduction. In T. Hoffmann & G. Trousdale (Eds.), *The Oxford handbook of construction grammar* (pp. 1–12). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195396683.013.0001
Hummel, K. (2006). *The traceback handbook* [unpublished manuscript]. Leipzig.
Ibbotson, P. (2020, August 31). The key to language is universal psychology, not universal grammar. *The Conversation*. https://theconversation.com/the-key-to-language-is-universal-psychology-not-universal-grammar-144220
Ibbotson, P. (2022). *Language acquisition: The basics*. Routledge. https://doi.org/10.4324/9781003156536
Ingram, D. (1989). *First language acquisition: Method, description and explanation*. Cambridge University Press.
Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198270126.001.0001
Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Cambridge University Press.
Johnston, J. R., & Slobin, D. I. (1979). The development of locative expressions in English, Italian, Serbo-Croatian and Turkish. *Journal of Child Language, 6*(3), 529–545. https://doi.org/10.1017/S030500090000252X
Karmiloff-Smith, A. (1998). Development itself is the key to understanding developmental disorders. *Trends in Cognitive Sciences*, *2*(10), 389–398. https://doi.org/10.1016/S1364-6613(98)01230-3
Kidd, E., & Garcia, R. (2022). How diverse is child language acquisition research? *First Language*, *42*(6), 703–735. https://doi.org/10.1177/01427237211066405

Kisilevsky, B. S., Hains, S. M. J., Brown, C. A., Lee, C. T., Cowperthwaite, B., Stutzman, S. S., Swansburg, M. L., Lee, K., Xie, X., Huang, H., Ye, H.-H., Zhang, K., & Wang, Z. (2009). Fetal sensitivity to properties of maternal speech and language. *Infant Behavior & Development*, *32*(1), 59–71. https://doi.org/10.1016/j.infbeh.2008.10.002

Koch, N. (2019). *Schemata im Erstspracherwerb. Eine Traceback-Studie für das Deutsche*. De Gruyter. https://doi.org/10.1515/9783110623857

Kol, S., Nir, B., & Wintner, S. (2014). Computational evaluation of the Traceback Method. *Journal of Child Language, 41*(1), 176–199. https://doi.org/10.1017/S0305000912000694

Kuhl, P. K., Andruski, J. E., Chistovitch, I. A., Chistovitch, L. A., Kozhevnikova, E. V., Ryskina, V. L., Stolyarova, E. I., Sundberg, U., & Lacerda, F. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science, 277*(5326), 684–686. https://doi.org/10.1126/science.277.5326.684

Lakoff, G., & Johnson, M. (1980). The metaphorical structure of the human conceptual system. *Cognitive Science, 4*(2), 195–208. https://doi.org/10.1016/S0364-0213(80)80017-6

Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago University Press. https://doi.org/10.7208/chicago/9780226471013.001.0001

Landau, B., & Jackendoff, R. (1993). "What" and "where" in spatial language and spatial cognition. *The Behavioral and Brain Sciences*, *16*(2), 217–238. https://doi.org/10.1017/S0140525X00029733

Langacker, R. W. (1987). *Foundations of cognitive grammar (Vol. 1): Theoretical prerequisites*. Stanford University Press.

Langacker, R. W. (2000) *Grammar and conceptualization*. De Gruyter. https://doi.org/10.1515/9783110800524

Langacker, R. W. (2008). *Cognitive grammar: A basic introduction*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195331967.001.0001

Langacker, R. W. (2009). A dynamic view of usage and language acquisition. *Cognitive Linguistics*, *20*(3), 627–640. https://doi.org/10.1515/COGL.2009.027

Langacker, R. W. (2013). *Essentials of cognitive grammar*. Oxford University Press.

Levine, S. C., Ratliff, K. R., Huttenlocher, J., & Cannon, J. (2012). Early puzzle play: A predictor of preschoolers' spatial transformation skill. *Developmental Psychology, 48*(2), 530–542. https://doi.org/10.1037/a0025913

Levinson, S. C. (1996). Language and space. *Annual Review of Anthropology*, *25*(1), 353–382. https://doi.org/10.1146/annurev.anthro.25.1.353

Levinson, S. C. (2003). *Space in language and cognition: Explorations in cognitive diversity*. Cambridge University Press. https://doi.org/10.1017/CBO9780511613609

Levinson, S. C., & Wilkins, D. P. (2006). The background to the study of the language of space. In S. C. Levinson & D. P. Wilkins (Eds.), *Grammars of space* (pp. 1–23). Cambridge University Press. https://doi.org/10.1017/CBO9780511486753.002

Lieven, E. (2010). Input and first language acquisition: Evaluating the role of frequency. *Lingua, 120*(11), 2546–2556. https://doi.org/10.1016/j.lingua.2010.06.005

Lieven, E., & Behrens, H. (2012). Dense Sampling. In E. Hoff (Ed.), *Research methods in child language* (pp. 226–239). Wiley. https://doi.org/10.1002/9781444344035.ch15

Lieven, E., Behrens, H., Speares, J., & Tomasello, M. (2003). Early syntactic creativity: A usage-based approach. *Journal of Child Language, 30*(2), 333–370. https://doi.org/10.1017/S0305000903005592

Lieven, E., & Goh, J. (n.d.-a). *CHILDES English MPI-EVA-Manchester Corpus: Project description*. Talkbank. https://childes.talkbank.org/access/Eng-UK/MPI-EVA-Manchester.html

Lieven, E., & Goh, J. (n.d.-b). *CHILDES English Thomas Corpus: Project description*. Talkbank. https://childes.talkbank.org/access/Eng-UK/Thomas.html
Lieven, E., Pine, J. M., & Baldwin, G. (1997). Lexically-based learning and early grammatical development. *Journal of Child Language, 24*(1), 187–219. doi:10.1017/S0305000996002930
Lieven, E., Salomo, D., & Tomasello, M. (2009). Two-year-old children's production of multiword utterances: A usage-based analysis. *Cognitive Linguistics, 20*(3). https://doi.org/10.1515/COGL.2009.022
Lieven, E., & Tomasello, M. (2008). Children's first language acquisition from a usage-based perspective. In P. Robinson & N. C. Ellis (Eds.), *Handbook of cognitive linguistics and second language acquisition* (pp. 168–196). Erlbaum. https://doi.org/10.4324/9780203938560-16
Lindblom, B., MacNeilage, P., & Studdert-Kennedy, M. (1984). Self-organizing processes and the explanation of phonological universals. In B. Butterworth, B. Comrie, & Ö. Dahl (Eds.), *Explanations for language universals* (pp. 181–203). Mouton. https://doi.org/10.1515/9783110868555.181
Lust, B. C. (2006). *Child language: acquisition and growth*. Cambridge University Press. https://doi.org/10.1017/cbo9780511803413
MacWhinney, B. (1996). The CHILDES System. *American Journal of Speech-Language Pathology*, *5*(1), 5–14. https://doi.org/10.1044/1058-0360.0501.05
MacWhinney, B. (2014). Item-based patterns in early syntactic development. In T. Herbst, H.-J. Schmid, & S. Faulhaber (Eds.), *Constructions, collocations, patterns* (pp. 33–69). De Gruyter. https://doi.org/10.1515/9783110356854.33
MacWhinney, B. (2024, October 10). *The CHILDES project: Tools for analyzing talk – electronic edition. Part 1: The CHAT transcription format*. Talkbank. http://childes.psy.cmu.edu/manuals/chat.pdf
MacWhinney, B. (2025, January 15). *The CHILDES project: Tools for analyzing talk – electronic edition. Part 2: The CLAN programs*. Talkbank. http://childes.psy.cmu.edu/manuals/clan.pdf
MacWhinney, B., & Snow, C. (1985). The child language data exchange system. *Journal of Child Language*, *12*(2), 271–295. https://doi.org/10.1017/S0305000900006449
MacWhinney, B., & Snow, C. (1990). The Child Language Data Exchange System: An update. *Journal of Child Language*, *17*(2), 457–472. https://doi.org/10.1017/S0305000900013866
Meints, K., Plunkett, K., Harris, P. L., & Dimmock, D. (2002). What is "on" and "under" for 15-, 18- and 24- month-olds? Typicality effects in early comprehension of spatial prepositions. *British Journal of Developmental Psychology, 20*(1), 113–130. https://doi.org/10.1348/026151002166352
Miller, G. A., & Johnson-Laird, P. N. (1976). *Language and perception*. Harvard University Press. https://doi.org/10.4159/harvard.9780674421288
Miller, H. E., Vlach, H. A., & Simmering, V. R. (2016). Producing spatial words is not enough: Understanding the relation between language and spatial cognition. *Child Development*, *88*(6), 1966–1982. https://doi.org/10.1111/cdev.12664
Mintz, T. H. (2003). Frequent frames as a cue for grammatical categories in child directed speech. *Cognition, 90*(1), 91–117. https://doi.org/10.1016/S0010-0277(03)00140-9
Miorelli, L. (2017). *The development of morpho-syntactic competence in Italian-speaking children: A usage-based approach* (Publication No. 27992219) [Doctoral dissertation, Northumbria University]. ProQuest Dissertations & Theses.

Morse, A. F., & Cangelosi, A. (2017). Why are there developmental stages in language learning? A developmental robotics model of language development. *Cognitive Science*, *41*(S1), 32–51. https://doi.org/10.1111/cogs.12390
Needham, A., & Baillargeon, R. (1993). Intuitions about support in 4.5-month-old infants. *Cognition*, *47*(2), 121–148. https://doi.org/10.1016/0010-0277(93)90002-D
Ochs, E. (1979). Transcription as theory. In E. Ochs & B. Schieffelin (Eds.), *Developmental pragmatics* (pp. 43–71). Academic Press.
O'Keefe, J., & Nadel, L. (1978). *The hippocampus as a cognitive map*. Oxford University Press.
Östman, J.-O., & Fied, M. (2004). Historical and intellectual background of construction grammar. In M. Fried & J.-O. Östman (Eds.), *Construction grammar in a cross-language perspective* (pp. 1–16). John Benjamins. https://doi.org/10.1075/cal.2.01ost
Parker, M. D., & Brorson, K. (2005). A comparative study between mean length of utterances in morphemes (MLUm) and mean length of utterances in words (MLUw). *First Language, 25*(3), 365–376. https://doi.org/10.1177/0142723705059114
Peters, A. M. (1977). Language learning strategies: Does the whole equal the sum of the parts? *Language, 53*(3), 560–573. https://doi.org/10.2307/413177
Peters, A. M. (1983). *The units of language acquisition*. Cambridge University Press.
Piaget, J., & Inhelder, B. (1956). *The child's conception of space*. Routledge.
Pine, J. M., & Lieven, E. (1993). Reanalysing rote-learned phrases: Individual differences in the transition to multi-word speech. *Journal of Child Language, 20*(3), 551–571. https://doi.org/10.1017/S0305000900008473
Pinker, S. (1984). *Language learnability and language development*. Harvard University Press.
Pinker, S. (1994). *The language instinct*. William Morrow.
Quick, A. E., & Hartmann, S. (2021). The building blocks of child bilingual code-mixing: A cross-corpus traceback approach. *Frontiers in Psychology, 12*, Article 682838. https://doi.org/10.3389/fpsyg.2021.682838
Quinn, P. Q. (1994). The categorization of above and below spatial relations by young infants. *Child Development*, *65*(1), 58–69. https://doi.org/10.2307/1131365
Quinn, P. Q., Adams, A., Kennedy, E., Shettler, L., & Wasnik, A. (2003). Development of an abstract category representation for the spatial relation between in 6- to 10-month old infants. *Developmental Psychology*, *39*(1), 151–163. https://doi.org/10.1037/0012-1649.39.1.151
Radden, G., & Dirven, R. (2007). *Cognitive English grammar*. John Benjamins Publishing. https://doi.org/10.1075/clip.2
Roseberry, S., Göksun, T., Hirsh-Pasek, K., & Golinkoff, R. M. (2012). Carving categories in a continuous world: Preverbal infants discriminate categorical changes before distance changes in dynamic events. *Cognition and Computation*, *12*(4), 231–251. https://doi.org/10.1080/13875868.2011.564338
Rowland, C. (2014). *Understanding child language acquisition*. Routledge. https://doi.org/10.4324/9780203776025
Rowland, C. F., Fletcher, S. L., & Freudenthal, D. (2008). How big is big enough? Assessing the reliability of data from naturalistic samples. In H. Behrens (Ed.), *Corpora in language acquisition research: History, methods, perspectives* (pp. 1–24). John Benjamins Publishing. https://doi.org/10.1075/tilar.6.04row
Sag, I., Boas, H. C., & Kay, P. (2012). Introducing sign-based construction grammar. In H. C. Boas & I. Sag (Eds.), *Sign-based construction grammar* (pp. 1–29). CSLI Publications.

Saxton, M. (1996). ‘No negative evidence’: What’s the problem?. *Proceedings of the Annual Meeting of the Berkeley Linguistics Society, 22*(1), 514–524. https://doi.org/10.3765/bls.v22i1.1324
Shatz, M. (2007). On the development of the field of language development. In E. Hoff & M. Shatz (Eds.), *Blackwell handbook of language development* (pp. 1–15). Blackwell Publishing. https://doi.org/10.1002/9780470757833.ch1
Shusterman, A., & Li, P. (2016). Frames of reference in spatial language acquisition. *Cognitive Psychology, 88*, 115–161. https://doi.org/10.1016/j.cogpsych.2016.06.001
Slobin, D. I. (2014). Before the beginning: The development of tools of the trade. *Journal of Child Language, 41*(S1), 1–17. https://doi.org/10.1017/S0305000914000166
Smoczynska, M. (2001). Studying Jan Baudouin de Courtenay’s Polish child language data 100 years later. In R. Finazzi & P. Tornaghi (Eds.), *Cinquant'anni di ricerche linguistiche: Problemi, risultati e prospettive per il terzo millennio* (pp. 591–610). Edizioni dell’ Orso.
Stern, C., & Stern, W. (1907). *Die Kindersprache.* Barth.
Stock, K., Jones, C. B., & Tenbrink, T. (2022). Speaking of location: A review of spatial language research. *Spatial Cognition and Computation*, *22*(3–4), 185–224. https://doi.org/10.1080/13875868.2022.2095275
Stoll, S., Abbot-Smith, K., & Lieven, E. (2009). Lexically restricted utterances in Russian, German, and English child-directed speech. *Cognitive Science*, *33*(1), 75–103. https://doi.org/10.1111/j.1551-6709.2008.01004.x
Svorou, S. (1994). *Grammar of space*. John Benjamins Publishing. https://doi.org/10.1075/tsl.25
Talmy, L. (1983). How language structures space. In H. L. Pick & L. P. Acredolo (Eds.), *Spatial orientation: Theory, research, and application* (pp. 225–282). Plenum.
Talmy, L. (2000). *Toward a cognitive semantics, volume 1: Concept structuring systems*. MIT Press. https://doi.org/10.7551/mitpress/6847.001.0001
Tardif, T., Fletcher, P., Liang, W., Zhang, Z., Kaciroti, N., & Marchman, V. A. (2008). Baby’s first 10 words. *Developmental Psychology*, *44*(4), 929–938. https://doi.org/10.1037/0012-1649.44.4.929
Theakston, A., & Lieven, E. (2017). Multiunit sequences in first language acquisition. *Topics in Cognitive Science, 9*(3), 588–603. https://doi.org/10.1111/tops.12268
Tomasello, M. (1987). Learning to use prepositions: A case study. *Journal of Child Language, 14*(1), 79–98. https://doi.org/10.1017/s0305000900012745
Tomasello, M. (1992). *First verbs: A case study of early grammatical development*. Cambridge University Press. https://doi.org/10.1017/CBO9780511527678
Tomasello, M. (2000). First steps toward a usage-based theory of language acquisition. *Cognitive Linguistics*, *11*(1–2), 61–82. https://doi.org/10.1515/cogl.2001.012
Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Harvard University Press. https://doi.org/10.2307/j.ctv26070v8
Tomasello, M. (2006). Acquiring linguistic constructions. In D. Kuhn, & R. S. Siegler (Eds.), *Handbook of child psychology: Cognition, perception, and language* (6th ed., pp. 255–298). Wiley.
Tomasello, M., & Brooks, P. J. (1999). Early syntactic development: A construction grammar approach. In M. D. Barrett (Ed.), *The development of language* (pp. 161–189). Psychology Press.
Tomasello, M., & Carpenter, M. (2007). Shared intentionality. *Developmental Science*, *10*(1), 121–125. https://doi.org/10.1111/j.1467-7687.2007.00573.x

Tomasello, M., & Stahl, D. (2004). Sampling children's spontaneous speech: How much is enough? *Journal of Child Language*, *31*(1), 101–121. https://doi.org/10.1017/S0305000903005944

Van der Hulst, H. (2023). The stages of language acquisition. In H. van der Hulst (Ed.), *A mind for language: An introduction to the innateness debate* (pp. 337–380). Cambridge University Press. https://doi.org/10.1017/9781108647182.010

Vihman, M. M. (1996). *Phonological development: The origins of language in the child*. Blackwell Publishing.

Vorster, J. (1984, July 9–13). *The first prepositions in Afrikaans: Order and semantic distinctions* [Paper presentation]. Third international congress for the study of child language, Austin, TX, United States.

Weist, R. M., Lymburner, N. L., Piotrowski, S., & Stoddard, J. L. (2000). Spatial complexity in children's language. *Perceptual and Motor Skills, 91*(2), 425–434. https://doi.org/10.2466/pms.2000.91.2.425

Wierzbicka, A. (1993). Why do we say in April, on Thursday, at 10 o'clock? In search of an explanation. *Studies in Language*, *17*(2), 437–454. https://doi.org/10.1075/sl.17.2.07wie

Wray, A. (2012). What do we (think we) know about formulaic language? An evaluation of the current state of play. *Annual Review of Applied Linguistics, 32*(1), 231–254. https://doi.org/10.1017/s026719051200013x

Wu, D., Li, H., & Degotardi, S. (2022). Early spatial language development and education: A scoping review. *SN Social Sciences, 2*, Article 20. https://doi.org/10.1007/s43545-021-00260-0

Yang, X., & Pan, Y. (2021). Spatial language of young children during block play in kindergartens in urban China. *Frontiers in Psychology*, *12*, Article 568638. https://doi.org/10.3389/fpsyg.2021.568638

Ziem, A., & Lasch, A. (2013). *Konstruktionsgrammatik: Konzepte und Grundlagen gebrauchsbasierter Ansätze*. De Gruyter. https://doi.org/10.1515/9783110295641

9. Appendix

The appendix includes the datasheets indicating whether the target utterances were derivable from the main corpora, and if so, how many operations were required to create the match. The respective tables for Eleanor (Table 34), Fraser (Table 35) and Thomas (Table 36) are attached on the following pages. First, however, some explanations regarding the shorthand labels of the columns are needed:

Nr .... utterance number in original transcript
Target Utterance .... child production as transcribed in the corpus
Traceback .... success (yes) or fail (no)
Operations .... number of operations required in total
SUB .... number of substitutions required
SUP .... number of superimpositions required
ADD .... number of juxtapositions required

Table 34. Target utterances and their derivability in the Eleanor-corpus

Nr	Target Utterance (in Test Corpus 020600a.cha)	Traceback	Operations	SUB	SUP	ADD
13	CHI: get in [?] the rabbit.	no (syntactic fail)	/	/	/	/
14	CHI: oh there you are.	yes	exact match	0	0	0
16	CHI: in there.	yes	exact match	0	0	0
51	CHI: look at my big stomach.	yes	2	1	1	0
62	CHI: he's got a small stomach.	yes	2	1	1	0
74	CHI: daddy_long_legs.	no (lexical fail)	/	/	/	/
106	CHI: where's dummy [?]?	yes	1	1	0	0
107	CHI: where's Jeannine?	yes	1	1	0	0
115	CHI: how do we fold it?	no (lexical fail)	/	/	/	/
141	CHI: it is there.	no (syntactic fail)	/	/	/	/
142	CHI: there.	yes	exact match	0	0	0
146	CHI: will it go in?	yes	1	1	0	0
147	CHI: where's it going?	yes	1	0	1	0
150	CHI: now he can go in.	yes	2	1	1	0
151	CHI: now all my necks [?] in.	no (syntactic fail)	/	/	/	/
155	CHI: climb in with daughter [?] baby.	no (lex + synt)	/	/	/	/
166	CHI: look at her.	yes	1	1	0	0
169	CHI: clean this space.	yes	1	1	0	0
172	CHI: jump.	yes	exact match	0	0	0
174	CHI: go and find my black book in the garden.	yes	4	3	0	1
176	CHI: &-um can't find my rubber anywhere.	yes	2	2	0	0
193	CHI: please come in.	yes	1	1	0	0
215	CHI: where's the fish fry?	no (lexical fail)	/	/	/	/
223	<did a &~Pol> [//] <are you just tall> [?], Polly.	no (lex +synt)	/	/	/	/
243	CHI: there's a nappie.	yes	1	1	0	0
245	CHI: &-um she's going to bed now.	yes	3	0	2	1
250	CHI: in her cot.	no (syntactic fail)	/	/	/	/

255	CHI: you hafta sleep there.	yes	2	2	0	0
281	CHI: where [>]?	yes	exact match	0	0	0
290	CHI: ah that's for this one.	yes	1	1	0	0
293	CHI: this is me washing clothes.	no (syntactic fail)	/	/	/	/
294	CHI: &-um [/] u@l Polly did a poo on it.	no (lexical fail)	/	/	/	/
297	CHI: in the washing machine.	no (lexical fail)	/	/	/	/
300	CHI: push fresh [?] washing.	yes	2	2	/	/
302	CHI: put it on the radiator@p.	yes	1	0	1	0
310	CHI: (be)cause I [/] I don't need a blanket on.	yes	2	1	0	1
323	CHI: push it.	yes	exact match	0	0	0
324	CHI: where's my washing gone?	yes	2	2	0	0
336	CHI: I'm [<] gonna [: going to] go back to my house.	yes	3	2	1	0
345	CHI: this is one holly [?].	no (syntactic fail)	/	/	/	/
352	CHI: stay in there.	yes	1	1	0	0
411	CHI: this one is her.	yes	1	1	0	0
417	CHI: &-um still in the bag.	yes	1	1	0	0
455	CHI: <this one does go> [//] this goes right [?].	yes	2	1	0	1
457	CHI: no not that.	yes	1	1	0	0
475	CHI: Mummy goes to there.	yes	2	2	0	0
485	CHI: let's go (.), lazy.	no (syntactic fail)	/	/	/	/
491	CHI: we're going now.	yes	1	1	0	0
499	CHI: we're watching all types there.	no (lexical fail)	/	/	/	/
503	CHI: can I go?	yes	exact match	0	0	0
505	CHI: can I get Polly out?	no (lexical fail)	/	/	/	/
519	CHI: there is that, Mummy.	yes	2	0	1	1
522	CHI: this is my pool [?].	no (syntactic fail)	/	/	/	/
524	CHI: Dalmations is in here.	no (lexical fail)	/	/	/	/
539	CHI: here comes my dinner mat.	no (lex + synt)	/	/	/	/
551	CHI: this is Dalmations <in the way> [?].	no (lex + synt)	/	/	/	/
560	CHI: sit there.	yes	exact match	0	0	0

590	CHI: they're in my bedroom [?] with Polly.	no (lex + synt)	/	/	/	/
628	CHI: I'm going to the shops.	yes	1	0	1	0
630	CHI: you stay here thank you.	yes	2	0	1	1
638	CHI: here.	yes	exact match	0	0	0
662	CHI: that's my big baby.	yes	1	0	1	0
665	CHI: &-um &-um I wanna play that.	yes	2	1	1	0
666	CHI: here you go [>].	yes	exact match	0	0	0
671	CHI: that's a hammock <upstanding> [?].	no (lexical fail)	/	/	/	/
674	CHI: that's a <book work> [?].	no (lexical fail)	/	/	/	/
718	CHI: I'm on it.	yes	1	1	0	0
756	CHI: and I see it there.	yes	3	1	1	1
798	CHI: open your mouth.	yes	exact match	0	0	0
802	CHI: <did um> [/] <did um> [//] shall I go and find my cat?	yes	3	2	1	0
803	CHI: &-uh &-uh <shall I> [/] <shall I go> [/] shall I go and find my cat and then stroke him?	no (syntactic fail)	/	/	/	/
826	CHI: look up in the bin.	no (syntactic fail)	/	/	/	/
834	CHI: let's go, Mummy.	yes	1	0	0	1
837	CHI: she's in the pushchair.	no (lexical fail)	/	/	/	/
845	CHI: here Polly.	no (lexical fail)	/	/	/	/
852	CHI: <get out> [=! whispers].	yes	exact match	0	0	0
864	CHI: <I don't wanna go in the lounge> [>].	yes	1	0	1	0
869	CHI: I wanna sit <here on the towel> [?] [>].	yes	2	1	1	0
987	CHI: and [<] she fell down.	yes	2	1	0	1
1069	CHI: &-um (be)cause I'm not a big girl yet.	yes	3	1	1	1
1072	CHI: where's Polly?	no (lexical fail)	/	/	/	/
1077	CHI: let's go.	yes	exact match	0	0	0
1078	CHI: you be the doctor <and (.) you be the hospital> [>].	yes	3	2	0	1
1118	CHI: where is the medicine?	yes	1	0	1	0
1139	CHI: there is the jigsaws [?].	yes	1	1	0	0
1140	CHI: there's this other one.	yes	2	1	1	0

1162	CHI: <shall I> [/] shall I pull this out?	yes	2	2	0	0
1185	CHI: that one.	yes	exact match	0	0	0
1234	CHI: they don't need that one.	yes	2	2	0	0
1277	CHI: little [?] one.	yes	exact match	0	0	0
1319	CHI: it's a boy at the wood.	no (syntactic fail)	/	/	/	/
1341	CHI: push me please.	yes	2	1	0	1
1356	CHI: &-um (.) a yellow piece.	yes	1	1	0	0
1405	CHI: this is that one.	yes	1	1	0	0
1439	CHI: here it is.	yes	1	0	1	0
1488	CHI: goes in here.	yes	1	1	0	0
1493	CHI: go away.	yes	exact match	0	0	0
1494	CHI: come back another day.	yes	1	1	0	0
1510	CHI: let's lie down for a bit.	yes	2	2	0	0
1517	CHI: go in another one.	no (syntactic fail)	/	/	/	/

Table 35. Target utterances and their derivability in the Fraser-corpus

Nr	Target Utterance (in Test Corpus Fraser 020601a.cha)	Traceback	Operations	SUB	SUP	ADD
78	CHI: <I xxx> [//] I dropped it. [+ PI]	yes	exact match	0	0	0
82	CHI: gonna [: going to] eat up those Quavers.	no (syntactic fail)	/	/	/	/
95	CHI: all the big books.	yes	1	/	1	/
123	CHI: book there.	no (syntactic fail)	/	/	/	/
146	CHI: where's a [*] crisp?	yes	1	0	1	0
170	CHI: what [//] where's my book?	yes	1	0	1	0
227	CHI: <what is> [?] that one.	yes	1	1	0	0
252	CHI: sliding on his tail.	no (lexical fail)	/	/	/	/
256	CHI: oh he's sliding down his tail.	no (lexical fail)	/	/	/	/
262	CHI: this mouse is running up the cat's black leg.	no (syntactic fail)	/	/	/	/
278	CHI: hiding behind the sofa [>].	yes	1	1	0	0
321	CHI: just (.) spinning.	no (lexical fail)	/	/	/	/

329	CHI: do like this.	yes	exact match	0	0	0
349	CHI: doing [=? going] like that.	yes	1	1	0	0
583	CHI: I want this one.	yes	exact match	0	0	0
591	CHI: in this book?	yes	1	0	1	0
595	CHI: there's no giraffe in here.	yes	2	1	0	1
612	CHI: there's one!	yes	exact match	0	0	0
623	CHI: I wanna botbot@f [= baby bottle] like that cow.	no (syntactic fail)	/	/	/	/
654	CHI: could you read this for me?	no (syntactic fail)	/	/	/	/
656	CHI: <I see> [<] a red big dragon.	no (syntactic fail)	/	/	/	/
659	CHI: Jack_in_the_box!	no (lexical fail)	/	/	/	/
664	CHI: that's Jwww [% brother] playing with blocks.	yes	2	2	0	0
677	CHI: and that's me.	yes	1	1	0	0
697	CHI: that's his tail.	yes	1	0	1	0
700	CHI: that blue dragon scared [*] that (.) bird.	no (syntactic fail)	/	/	/	/
707	CHI: that's Daddy climbing (.) up in the air.	yes	2	1	1	0
710	CHI: he's flying.	yes	1	1	0	0
714	CHI: it's not here.	no (syntactic fail)	/	/	/	/
715	CHI: that's you.	yes	exact match	0	0	0
729	CHI: red (.) big (.) dragon.	no (syntactic fail)	/	/	/	/
778	CHI: I wanna go downstairs and have a botbot@f!	yes	3	2	0	1
783	CHI: (be)cause &-um &-um I wanna go downstairs.	yes	2	0	1	1
784	CHI: can you open the door?	yes	2	2	0	0
795	CHI: and a big squeeze.	yes	1	0	1	0
800	CHI: and a big kiss.	yes	1	0	1	0
806	CHI: and a big hug.	yes	exact match	0	0	0
887	CHI: I [/] <I want> [/] I wanna open the door.	yes	2	0	2	0
894	CHI: I can open it.	yes	1	0	1	0
900	CHI: it's dark in here.	yes	exact match	0	0	0
925	CHI: that's a funny noise.	yes	exact match	0	0	0
938	CHI: but it can go?	yes	2	0	1	1

956	CHI: and I put toys <in my> [//] in [/] in my &~b [//] bag.	yes	4	1	2	1
964	CHI: that bear [?] monster.	no (syntactic fail)	/	/	/	/
1018	CHI: &~wh [//] <where's Dad> [//] where's Daddy?	yes	exact match	0	0	0
1020	CHI: &~wh [//] where's Jwww [% brother] and Daddy?	yes	2	2	0	0
1022	CHI: where's Jwww [% brother]?	yes	exact match	0	0	0
1026	CHI: downstairs.	yes	exact match	0	0	0
1148	CHI: I'm building a tower.	yes	exact match	0	0	0
1152	CHI: there's another block over there.	yes	1	1	0	0
1157	CHI: there's one over there.	yes	1	1	0	0
1160	CHI: and there's another book.	yes	1	1	0	0
1169	CHI: look at this book.	yes	1	1	0	0
1181	CHI: is that a hippopotamus?	no (lexical fail)	/	/	/	/
1212	CHI: on [?] the grass.	yes	exact match	0	0	0
1217	CHI: he dropped it.	yes	1	1	0	0
1219	CHI: he dropped it on the ground.	yes	2	2	0	0
1267	CHI: don't touch it.	yes	1	0	1	0
1301	CHI: is the car gone?	yes	1	1	0	0
1311	CHI: I hurt on those &~c [//] track.	no (lexical fail)	/	/	/	/
1317	CHI: I wanna go to [//] open the gate.	yes	exact match	0	0	0
1331	CHI: do that.	yes	exact match	0	0	0
1334	CHI: that isn't Jwww's [% brother] picture.	no (syntactic fail)	/	/	/	/
1354	CHI: and I draw a rocket like this one.	no (lex + synt)	/	/	/	/
1445	CHI: and [/] and there was [*] crabs on the floor.	yes	3	1	1	1
1469	CHI: there's another crab.	yes	1	1	0	0
1509	CHI: he's flying really high.	yes	2	2	0	0
1523	CHI: <there's a> [//] that [/] that [/] <that's &-um> [//] (.) that [/] that's a bottle.	yes	1	1	0	0
1545	CHI: they're playing football in the park.	yes	2	1	0	1
1652	CHI: it's right in Jwww's [% brother] bedroom.	yes	1	1	0	0
1654	CHI: I show you where it is.	no (syntactic fail)	/	/	/	/

1669	CHI: it's not in my hand.	yes	1	1	0	0
1672	CHI: is it in here?	yes	1	0	1	0
1687	CHI: is that Liz [: Liz's] [*] bedroom?	yes	2	2	0	0
1704	CHI: it [//] it's not gonna [: going to] be in my lamp.	no (syntactic fail)	/	/	/	/
1708	CHI: there!	yes	exact match	0	0	0
1709	CHI: there it is!	yes	exact match	0	0	0
1731	CHI: is that a crocodile?	yes	1	0	1	0
1736	CHI: that banana in a drink?	yes	2	2	0	0
1768	CHI: what's that?	yes	exact match	0	0	0
1770	CHI: and this?	yes	exact match	0	0	0
1772	CHI: what's this?	yes	exact match	0	0	0
1775	CHI: <and what> [/] and what's this?	yes	exact match	0	0	0
1787	CHI: yeah it's got a big long [/] long neck.	yes	2	1	0	1
1796	CHI: is that a [*] (i)guana?	yes	1	1	0	0
1809	CHI: that [/] that's not a hippopotamus [= actually says hitapotamus].	no (lexical fail)	/	/	/	/
1832	CHI: and what's those?	yes	1	1	/	/
1854	CHI: but what's this?	yes	exact match	0	0	0
1857	CHI: but [<] what's that?	yes	1	1	0	0
1866	CHI: that [//] is it a robin?	yes	1	0	1	0
1871	CHI: what's this one?	yes	exact match	0	0	0
1886	CHI: is [/] is that icecream <and a> [/] and a icelolly?	yes	2	2	0	0
1912	CHI: is it [//] that a Yeti?	yes	1	0	1	0
1918	CHI: &-um (.) there's no painting on that one.	yes	2	1	0	1

Table 36. Target utterances and their derivability in the Thomas-corpus

Nr	Target Utterance (in Test Corpus Thomas 020602.cha)	Traceback	Operations	SUB	SUP	ADD
11	CHI: no back [?] it.	no (syntactic fail)	/	/	/	/
13	CHI: back fall down.	no (syntactic fail)	/	/	/	/
26	CHI: a big bus.	yes	1	1	0	0

38	CHI: oh dear lid come off.	yes	2	1	0	1
62	CHI: stop &~de train.	yes	1	1	0	0
68	CHI: a big fire.	yes	1	0	1	0
73	CHI: like that.	yes	exact match	0	0	0
79	CHI: a@p &~na train like that.	yes	1	1	0	0
104	CHI: none there.	yes	exact match	0	0	0
121	CHI: not a@p (.) a@p back fall down.	no (syntactic fail)	/	/	/	/
138	CHI: no up.	yes	exact match	0	0	0
218	CHI: that's better.	yes	exact match	0	0	0
257	CHI: what's this?	yes	exact match	0	0	0
271	CHI: this all a@p wet.	yes	2	2	0	0
356	CHI: &~de window fall down.	yes	1	1	0	0
365	CHI: not get in.	no (syntactic fail)	/	/	/	/
391	CHI: open it.	yes	exact match	0	0	0
396	CHI: 0 [=! babble] not get in there.	yes	2	1	0	1
399	CHI: oh racing car go in there.	yes	2	2	0	0
401	CHI: no going in there.	yes	2	1	0	1
409	CHI: here you are, racing car.	yes	1	1	0	0
410	CHI: racing car going in there.	yes	2	2	0	0
418	CHI: going backwards.	yes	exact match	0	0	0
419	CHI: going backwards, Mummy.	yes	1	0	0	1
423	CHI: train go in there.	yes	1	1	0	0
425	CHI: look coming in there.	no (syntactic fail)	/	/	/	/
427	CHI: coming in there.	yes	1	1	0	0
450	CHI: behind &~de board [?].	yes	1	1	0	0
463	CHI: music come from?	no (syntactic fail)	/	/	/	/
483	CHI: choo choo tunnel.	yes	1	1	0	0
490	CHI: choo choo tunnel please.	yes	2	1	0	1
515	CHI: not go.	yes	1	1	0	0
523	CHI: no box.	yes	1	1	0	0

525	CHI: a@p box.	yes	exact match	0	0	0
534	CHI: fit in there.	yes	exact match	0	0	0
535	CHI: fit in. [+ SR]	yes	exact match	0	0	0
548	CHI: a@p fit.	yes	exact match	0	0	0
560	CHI: no fit in &~da box.	yes	1	0	0	1
565	CHI: no <fit in> [>].	yes	1	0	1	0
571	CHI: coming from.	yes	exact match	0	0	0
595	CHI: <a@p> [<] a@p big tunnel.	yes	1	1	0	0
599	CHI: &~na big tunnel. [+ SR]	yes	1	1	0	0
608	CHI: coming &~da train.	no (syntactic fail)	/	/	/	/
609	CHI: coming. [+ SR]	yes	exact match	0	0	0
614	CHI: coming in there.	yes	1	1	0	0
615	CHI: coming in. [+ SR]	yes	1	0	1	0
617	CHI: no people on it.	yes	2	1	0	1
622	CHI: no people train.	no (syntactic fail)	/	/	/	/
624	CHI: a station.	yes	1	1	0	0
627	CHI: here you are, train.	yes	1	0	0	1
634	CHI: people coming now.	yes	1	1	0	0
644	CHI: bus station.	yes	1	1	0	0
653	CHI: bus station now.	yes	2	1	0	1
655	CHI: bus stopping.	yes	1	0	1	0
662	CHI: a@p moving.	yes	exact match	0	0	0
688	CHI: platform. [+ I]	no (lexical fail)	/	/	/	/
696	CHI: a train station.	yes	1	1	0	0
762	CHI: fit a@p train.	yes	1	1	0	0
763	CHI: train come.	yes	1	1	0	0
772	CHI: a@p too big.	yes	exact match	0	0	0
774	CHI: Mummy too big.	yes	1	0	0	1
776	CHI: oh where &~da bus?	yes	1	0	1	0
783	CHI: yellow truck come by.	no (syntactic fail)	/	/	/	/

785	CHI: truck going backwards.	yes	2	1	1	0
845	CHI: no fall down.	yes	1	1	0	0
851	CHI: what's this?	yes	exact match	0	0	0
869	CHI: too big a@p bus.	no (syntactic fail)	/	/	/	/
871	CHI: bus too big.	no (syntactic fail)	/	/	/	/
872	CHI: too big a@p &~le bus.	no (syntactic fail)	/	/	/	/
874	CHI: too big &~de trains.	no (syntactic fail)	/	/	/	/
879	CHI: not get in.	no (syntactic fail)	/	/	/	/
886	CHI: no ride &~da bus.	yes	2	1	1	0
913	CHI: out_of the way.	yes	exact match	0	0	0
927	CHI: a@p back fall.	no (syntactic fail)	/	/	/	/
966	CHI: this one, Mummy.	yes	1	0	0	1
967	CHI: that one.	yes	exact match	0	0	0
970	CHI: here you are, Mummy.	yes	exact match	0	0	0
1066	CHI: a@p &~de supermarket.	yes	exact match	0	0	0
1142	CHI: no that one.	yes	1	1	0	0
1199	CHI: <Mummy here> [?].	yes	exact match	0	0	0
1225	CHI: &~sa back fall.	no (syntactic fail)	/	/	/	/
1232	CHI: &~isa back.	yes	exact match	0	0	0
1233	CHI: big fall.	yes	1	1	0	0
1235	CHI: oh train fall.	yes	1	1	0	0
1236	CHI: more train fall.	no (syntactic fail)	/	/	/	/
1249	CHI: what in there?	yes	exact match	0	0	0
1252	CHI: a@p there.	yes	exact match	0	0	0
1267	CHI: no back.	yes	1	1	0	0
1277	CHI: no hands down [?].	no (syntactic fail)	/	/	/	/
1278	CHI: &~ne hand on.	yes	1	1	0	0
1292	CHI: from home.	yes	1	1	0	0
1327	CHI: a train jump.	yes	1	1	0	0
1381	CHI: big willie.	yes	exact match	0	0	0

Abstract

Over the last decades, the field of developmental linguistics has been marked by disagreement concerning the factors enabling and influencing child language acquisition (e.g. Chomsky 1980; Pinker, 1994; Tomasello, 2003; Goldberg, 2006). Much usage-based research has demonstrated that children's first complex constructions are rote-learned repetitions or mildly modified variations of prior strings and thus either frozen or schematic in nature. As linguistic development has crucially been linked to literacy, determining the cognitive mechanisms by which young learners effortlessly acquire a complex system such as human language, matters to scholars of all theoretical backgrounds. Early spatial language, in turn, is widely associated with spatial cognition and mathematical skills. At their intersection, the acquisition of constructions featuring representations of space, such as locations, shapes, dimensions or transformations, deserves to be explored. This present study applies the traceback method to spatial language constructions by relating children's early multiunit productions to previous discourse. Data was collected from the Eleanor-, Fraser- and Thomas-corpus composed by Lieven and Goh (n.d.-a, n.d.-b), available for further investigations on the child language archive CHILDES (MacWhinney & Snow, 1985). The findings suggest that between 70.0 and 79.6 percent of the children's spatial constructions were successfully built from precedents. While lower mean length of utterance (MLU) tracebacks required no or just one operation to arrive at a match, the number of multi-operation tracebacks increases with MLU. Spatial language appears to be realised predominantly in frame architectures and only infrequently occupied an open slot in such schemas. This confirms that spatial language learning, like construction learning on the whole, is initially characterised by a high degree of specificity and constrained creativity; with growing linguistic experience, learners evidently generalise across instantiations to create schemas with open slots and ultimately fully abstract patterns.

Zusammenfassung

In den letzten Jahrzehnten war das Forschungsfeld der Sprachentwicklung von Uneinigkeiten bezüglich jener Faktoren, die es Kindern ermöglichen Sprache zur erlernen, geprägt (siehe Chomsky 1980; Pinker, 1994; Tomasello, 2003; Goldberg, 2006). Gebrauchsbasierte Studien zeigten, dass erste komplexere Konstruktionen entweder die Gestalt von Eins-zu-Eins Wiederholungen oder minimalen Abwandlungen von bereits produzierten Strukturen annehmen. Sie wurden unter den Begriffen *fixed strings* („verfestigte Mehrwortäußerungen“) und *item-based schemas* („teilweise verfestigte Schemata“) diskutiert. Da frühe Sprachkenntnisse vermehrt mit Lese- und Schreibfähigkeit im weiteren Entwicklungsverlauf in Verbindung gebracht wurden, ist es Sprachwissenschaftler*innen unterschiedlicher theoretischer Überzeugungen ein Anliegen, herauszufinden, wie Kinder ohne ersichtliche Schwierigkeiten das komplexe System Sprache meistern. Für *spatial language* („räumliche Sprache”) wurde belegt, dass sie eng mit Raumkognition und mathematischen Fähigkeiten verbunden ist. Der Überschneidung der beiden Gebiete, nämlich dem Erwerb von Konstruktionen mit sprachlichen Vermittlern von Räumlichkeit, beispielsweise Orten, Formen, Dimensionen oder Transformationen, wurde bisher zu wenig Beachtung geschenkt. Die vorliegende Traceback-Studie versucht zu zeigen, dass sich *spatial language* Konstruktionen aus bereits wahrgenommenem Input und erworbenen Chunks zusammensetzen lassen. Zur Durchführung wurden das Eleanor-, Fraser- und Thomas-Korpus von Lieven und Goh (n.d.-a, n.d.-b) herangezogen, die auf der Datenbank CHILDES (MacWhinney & Snow, 1985) frei zugänglich vorliegen. 70 bis 79.6 Prozent aller Produktionen konnten so mithilfe vorausgehender Äußerungen als Bausteine rekonstruiert werden. Während das Kind mit der niedrigsten *mean length of utterance* (MLU; „mittlere Äußerungslänge“) die meisten exakten lexikalischen Übereinstimmungen aufwies, stieg die Anzahl der Konstruktionen, die Mehrfach-Operationen zur Ableitung benötigten, mit zunehmender MLU. *Spatial language* erscheint vor allem in Rahmenelementen und besetzt nur selten einen variablen Slot. Dies bestätigt, dass der Erwerb räumlicher Sprache, wie Konstruktionslernen im Allgemeinen, zu Beginn auf teilspezifischen Schemata aufbaut und damit nur begingt kreativ ist. Durch sprachliche Erfahrung kann das Kind später vollständig abstrakte Muster abstrahieren.