Lexical Enrichment and Structural Complexity in the Acquisition of English as a First Language

Proroković, Jakov

SVEUČILIŠTE U ZADRU

POSLIJEDIPLOMSKI SVEUČILIŠNI STUDIJ
HUMANISTIČKE ZNANOSTI

Jakov Proroković

LEXICAL ENRICHMENT AND STRUCTURAL COMPLEXITY IN THE ACQUISITION OF ENGLISH AS A FIRST LANGUAGE: DITRANSITIVE, CAUSATIVE AND RELATIVE CONSTRUCTIONS

Doktorski rad

Zadar, 2024.

SVEUČILIŠTE U ZADRU

POSLIJEDIPLOMSKI SVEUČILIŠNI STUDIJ

HUMANISTIČKE ZNANOSTI

Jakov Proroković

LEXICAL ENRICHMENT AND STRUCTURAL COMPLEXITY IN THE ACQUISITION OF ENGLISH AS A FIRST LANGUAGE: DITRANSITIVE, CAUSATIVE AND RELATIVE CONSTRUCTIONS

Doktorski rad

Mentor

izv. prof. dr. sc. Marco Angster

Zadar, 2024.

SVEUČILIŠTE U ZADRU

TEMELJNA DOKUMENTACIJSKA KARTICA

I. Autor i studij

Ime i prezime: Jakov Proroković

Naziv studijskog programa: Poslijediplomski sveučilišni studij Humanističke znanosti

Mentor/Mentorica: izv. prof. dr. sc. Marco Angster

Datum obrane: 11. listopada 2024.

Znanstveno područje i polje u kojem je postignut doktorat znanosti: humanističke znanosti, filologija

II. Doktorski rad

Naslov: Lexical Enrichment and Structural Complexity in the Acquisition of English as a First Language: Ditransitive, Causative and Relative Constructions

UDK oznaka: 811.111'232

Broj stranica: 265

Broj slika/tablica: 21/47

Broj bilježaka: 106

Broj korištenih bibliografskih jedinica i izvora: 396

Broj priloga: 2

Jezik rada: Engleski

III. Stručna povjerenstva

Stručno povjerenstvo za ocjenu doktorskog rada:

1. doc. dr. sc. Frane Malenica, predsjednik/predsjednica
2. prof. dr. sc. Marijan Palmović, član/ica
3. izv. prof. dr. sc. Mojca Kompara Lukančič, član/ica

Stručno povjerenstvo za obranu doktorskog rada:

1. doc. dr. sc. Frane Malenica, predsjednik/predsjednica
2. prof. dr. sc. Marijan Palmović, član/članica
3. izv. prof. dr. sc. Mojca Kompara Lukančič, član/članica

UNIVERSITY OF ZADAR

BASIC DOCUMENTATION CARD

I. Author and study

Name and surname: Jakov Proroković

Name of the study programme: Postgraduate doctoral study programme in Humanities

Mentor: Associate Professor Marco Angster, PhD

Date of the defence: 11 October 2024

Scientific area and field in which the PhD is obtained: Humanities, Philology

II. Doctoral dissertation

Title: Lexical Enrichment and Structural Complexity in the Acquisition of English as a First Language: Ditransitive, Causative and Relative Constructions

UDC mark: 811.111'232

Number of pages: 265

Number of figures/tables: 21/47

Number of notes: 106

Number of used bibliographic units and sources: 396

Number of appendices: 2

Language of the doctoral dissertation: English

III. Expert committees

Expert committee for the evaluation of the doctoral dissertation:

1. Assistant Professor Frane Malenica, PhD, chair
2. Full Professor Marijan Palmović, PhD, member
3. Associate Professor Mojca Kompara Lukančič, PhD, member

Expert committee for the defence of the doctoral dissertation:

1. Assistant Professor Frane Malenica, PhD, chair
2. Full Professor Marijan Palmović, PhD, member
3. Associate Professor Mojca Kompara Lukančič, PhD, member

Izjava o akademskoj čestitosti

Ja, **Jakov Proroković,** ovime izjavljujem da je moj **doktorski** rad pod naslovom **Lexical Enrichment and Structural Complexity in the Acquisition of English as a First Language: Ditransitive, Causative and Relative Constructions** rezultat mojega vlastitog rada, da se temelji na mojim istraživanjima te da se oslanja na izvore i radove navedene u bilješkama i popisu literature. Ni jedan dio mojega rada nije napisan na nedopušten način, odnosno nije prepisan iz necitiranih radova i ne krši bilo čija autorska prava.

Izjavljujem da ni jedan dio ovoga rada nije iskorišten u kojem drugom radu pri bilo kojoj drugoj visokoškolskoj, znanstvenoj, obrazovnoj ili inoj ustanovi.

Sadržaj mojega rada u potpunosti odgovara sadržaju obranjenoga i nakon obrane uređenoga rada.

Zadar, 18. listopada 2024.

In Appreciation

I would like to express my heartfelt gratitude to my family—my wife, my parents, and my friends—who stood by me throughout the journey of creating this dissertation. Their patience and understanding during the more challenging periods of this work meant the world to me. I am also deeply thankful to my Department, which has been a steadfast source of support since I embarked on this campaign, to my former professors who inspired me long before I considered pursuing a PhD, and, finally, to my mentor, whose expertise and guidance were invaluable throughout this process.

To my baby boy!

Project Acknowledgment

This work has been realised under the institutional project *Multilingualism: between theory and empirical knowledge (ViTE – Višejezičnost: između teorije i empirije)* (IP-01-2023-14), fully financed by the University of Zadar.

Table of contents

List of tables

List of figures

Abbreviations

N – bare nominal (no modifiers)
MOD-N – nominal with modifiers
NP – any noun phrase (N or MOD-N or PRO)
PRO – pronoun
ProDem - demonstrative pronoun
ProPers – personal pronoun
V – verb
C – clause
REL – relativizer
INAN - inanimate objects
ANc – common nouns (teacher)
ANp – proper nouns (Bill)
ANt – toys (animals)
ANb – body-parts (hands, legs, etc.)
Inherent animates – ANc + ANp
Contextual animates – ANt + ANb
MLU – mean length of utterance
LD – lexical density
LS1 – lexical sophistication
CVS1 – corrected verb sophistication
NDW – number of different words
NDW-ER50 – NDW (expected random 50)
TTR – type-token ratio
MSTTR-50 – mean segmental TTR
CTTR - corrected TTR
RTTR – root TTR
SVV1 – squared verb variation

1. INTRODUCTION: CORPUS AND COGNITIVE LINGUISTICS

The main objective of this study is to explore how individuals of varying age groups (0-3, 4-6, and 18+) exhibit lexical and structural similarities within targeted linguistic constructions (ditransitive, periphrastic causative, and relative), using data from the CHILDES corpora, specifically focusing on English-speaking children. The field of language acquisition and development is immense in its investigative scope (both in the number of research issues it addresses and the number of studies conducted thus far), which is why the following subsections of the introduction serve to establish a general introduction to the topic, but also to provide an overview of theoretical frameworks deemed relevant in the context of this particular research design.

1.1. General introduction

There are several keywords that recurringly appear throughout this thesis that need to be addressed from the onset. One such keyword is the term 'construction', and its general and pattern-specific implications are discussed in the chapters to follow, but for the moment, it is important to recognize that specific combinations of words manifesting as stable linguistic patterns[1] can be treated as units in their own right, and this holds true in the context of language acquisition and development (cf. Goldberg & Casenhiser, 2008; Tomasello et al., 2008; Behrens, 2021). The terms 'acquisition' and 'development' are other keywords central to this work, and though they can often be used interchangeably, the distinction between the two is in the degree to which they emphasize the capacity for language and the mastery over it; the former emphasizes the subconscious process by which individuals learn their first language in a predictable sequence over a certain time period, while the latter encompasses a broader concept that refers to the ongoing refinement and expansion of language skills. It is often claimed that language acquisition is essentially finished by the age of 4, whereas the process of language

[1] Much like the term 'construction', the term 'pattern' is used frequently throughout the thesis, and though it may appear that the two are sometimes used rather interchangeably, the latter is primarily used to refer to specific lexical and syntactic arrangements within the observed constructions, such as the use of nouns, pronouns and even verbs forms in specific construction slots. For example, the periphrastic causative pattern can alternate depending on whether the embedded verb manifests a full infinitive (e.g. *she forced him to jump*) or bare infinitive form (e.g. *she made him jump*).

refinement continues, with significant improvements in articulation, vocabulary, sentence formation, and overall communicative abilities, especially in preschool and early school age (Hoff, 2013, p. 7).

Much like the interrelationship between the two terms paves the way for the discussion on what exactly falls into the realm of 'capacity', and what into the realm of 'refinement', the term construction embodies the continuum between unity and complexity. At the same time, it represents a unit of stable meaning that requires of the learner to comprehend the grammar behind it, while allowing for refinements with the progression of age and linguistic ability. The way in which children learn to use constructions in their language and how this relates to their understanding of word meanings is thus essential for properly outlining the acquisitional and developmental trajectory of language. Though it seems straightforward to say that children learn constructions by observing and internalizing how words and phrases fit together to create meaningful expressions, there is an intricate language development trajectory to keep track of between birth and the point at which they are able to produce and comprehend sentences correctly. According to what we know thus far, children initially combine words to express salient information and contribute something new, but their understanding of language constructions is not yet adult-like. They learn to identify language constructions by paying attention to how adults use words and constructions in context, and by categorizing events and their parts. One idea that has been particularly prolific in relation to the subsequent emergence of constructions is that of 'verb islands' (Tomasello, 1992) – the idea that children's early use of verbs is limited to a small number of constructions, often just one, and that their understanding of syntactic constructions for different event-types develops over time. Children need to attend to how adults use words and constructions in order to identify the constructions and the word meanings within them. Children first learn to map some words onto parts of events before they grasp the syntactic constructions appropriate to the encoding of different event-types (see Clark, 2009, pp. 167-174).[2] As they learn to use more adult-like constructions, their categorizations of event-types becomes more consistent.

[2] A comprehensive introduction into the trajectory of language acquisition is provided by Eve V. Clark (2009). Clark leans towards usage-based accounts in her overview, emphasizing the role of both general cognition and language input.

In terms of language input that children hear, the studies so far suggest that children hear 5,000–7,000 utterances per day,[3] only 15% (approx.) of which exhibit the canonical SVO form,[4] while the rest of the speech consists of the following: more than 20% are not full sentences (but fragmented speech in the form of noun phrases or prepositional phrases), 25-30% are questions, approximately 25% are imperatives and structures centred around a copula (Tomasello, 2005, p. 111). As far as the output is concerned, it is difficult to provide the exact temporal points in language development (since there are considerable differences between children themselves), but the syntactic milestones have been roughly noted by the researchers. In approximate terms, children start to produce single-word utterances as they turn 1, while the word combinations (or two-word utterances) where one item is usually represented by a determiner or a preposition, and the other is a content word, start to emerge after 18 months of age. The 'item-based' constructions[5] emerge after they turn two, while the more adult like constructions that have a certain level of abstractness to them, have been observed in 3 year olds and later (see Tomasello, 2005; also see Clark, 2009, p. 169). In other words, children start to exhibit the capacity for lexical partition of events as they turn 18 months of age, but the specification of participant roles is made evident only as they start to use item-based constructions.[6] According to Tomasello (1988, p. 80), 24 months of age is also the point at which the child begins to participate in "true conversation" by building upon the conversation at hand, or at least making remarks that share the topic that is being discussed. The greatest degree of syntactic capacity can be claimed once the child starts to combine both the specification of participants included

[3] In a notable study, Hart and Risley (2003) noted that the number of different words English-speaking children hear in early childhood greatly differs with respect to their parents' economic background: on average, a child in a family on welfare heard 616 words per hour, a child from a working-class family heard 1,251 words per hour, and a child in a family of working professionals was exposed to 2,153 words per hour.

[4] These include transitive utterances of various types that are typical of the English language, and the pronoun takes on the role of the subject in more than 80% of such cases (Tomasello, 2005, p. 111).

[5] The term implies that the linguistic abilities of young children primarily rely on individual words and phrases rather than larger systematic patterns, meaning that their early language development and understanding clings to specific, concrete, elements as opposed to system-wide syntactic schemas (Tomasello, 2000b). Although this has been initially posited for (and mostly observed on) specific verbs around which constructions are built, the term 'item-based' has been used in slightly more encompassing manner than 'verb-island', including the possibility of a wider range of central elements as potential construction-carriers.

[6] The development of attention skills basically underpins the linguistic production of children. The non-linguistic precursor to this level of language complexity in production is the ability of a child for joint attention (when they are able to coordinate their attention to objects and people at the same time), which emerges around 6 months of age (and starts to be linguistically demonstrated at 1), while the evolution of attention skills to assume a shared focus on a particular event/activity with their interlocutor emerges at 18 months of age (see Tomasello, 1988).

in the event described and the understanding/categorization of the referred to event types, which happens from age 3 onwards.

Children's early use of verbs is limited to specific nouns and simple constructions, but they gradually learn to use more elaborate constructions and combine verbs with different nouns and arguments.[7] They learn construction-types in relation to specific lexical items and they attend closely to the speech directed to them in picking up information on the constructions represented and the corresponding verbs used in them. Also, their understanding of semantic roles may be more restricted than that of adults, but they learn to treat agents, locations, and instruments of different actions alike. Two things have been observed and frequently asserted by the researchers in the area: (1) general-purpose verbs found across many languages are the ones that tend to be acquired early and reproduced frequently (Edwards & Goodwin, 1986; Theakston, 2004; Goldberg, 2013b; Ambridge et al., 2015b), and (2) there has been significant evidence that early children's language use, especially when it comes to specific constructions, centres around smaller number of items/verbs (Bloom, Miller, & Hood 1975; Pine, Lieven, & Rowland 1998).

One possible explanation for how children initially learn verbs is that they only learn them in the specific constructions they have heard them used in. Tomasello (2011, pp. 253-254) argued that this is especially true for the earliest stages of development, but not true for the entirety of early language development – given that the research has shown that children, as they get older, tend to make overgeneralisation errors. Some of the errors registered by researchers occurred already at 3 years of age (see Bowerman, 1982, 1988), when children applied a verb to a construction in which it does not typically belong (e.g. *Don't giggle me*). Another argument that has also been disproved to a certain point is that adults frequently correct these errors, and the linguistic environment imposes constrains on children's overgeneralisation tendencies. Tomasello (ibid.) claimed that these simple explanations are to be rejected, leaving us with three key propositions that researchers have put forward to explain how children learn

[7] Although the frequency effect is expected to diminish with age, there have been propositions of its relevance going beyond the earliest stages of language development; a study by Huttenlocher et al. (2002) done on 4-year olds found correlations between parent input and certain aspects of syntactic development, where the growth of particular forms was found to be in relation to parent's variation in speech (though the authors suggested that such relations can only be observed during the development, i.e. they should cease once the child has mastered the target form in full. Additionally, in her experiment with a semi-artificial language, Wonnacott (2011) found children older than 5 to be sensitive to linguistic distributions. Moreover, the lexically-based patterns of use were actually acquired by children in absence of any cues that would be otherwise available in non-controlled setting, meaning that the frequency effect is further isolated as a significant driver of language development.

which constructions best accommodate certain verbs: (1) there are specific semantic constraints that apply to certain English constructions and verbs (see Pinker, 1989), which children learn at some point, although it is not completely clear how; (2) the more frequently children hear a verb used in a particular construction, the less likely it will be for them to use it in a novel construction (Clark, 1987; Bates & MacWhinney, 1989 ; Braine & Brooks, 1995 and others) and (3) if children hear a verb used in a construction that serves the same purpose as a possible generalisation, they may recognize the non-canonical nature of that particular generalisation.[8]

This thesis does not necessarily put the focus entirely on verbs, but rather on the syntactic patterns within constructions along with distributional data pertaining to the lexical items in those construction, whether nouns or verbs. Nevertheless, similar logic can be applied to both nouns and constructions, that is, it is possible to rephrase the Tomasello's observation to encapsulate not only verbs, but nouns as well, if not even constructions in their entirety (as they represent and behave like items in their own respect). Aside from verbs, there may be constraints that children learn concerning the use of nouns and patterns, the frequency of these may facilitate the acquisitional process, and they may avoid using certain items and patterns which would otherwise appear sound to them – but due to lack thereof in the input – resort to those more frequent, even when they do not necessarily represent the "easier" route. In this study, most of the data is interpreted through the lens of frequency and inter-group correspondence,[9] but this does not mean that the interpretation of the results should solely be restricted to the most frequent items in the input and their relevance; that is, it takes into account structures and items that are overrepresented, as well as those that are underrepresented or marked.[10] A quote by Ibbotson and Tomasello (2009, p. 80) wonderfully outlines the way in

[8] In some sense, the last claim implies that input trumps generalisation; e.g. a child might expect a construction as *He danced his friend* but after hearing *He made his friend dance*, infer that the verb 'dance' does not occur in transitive constructions – particularly in view of the fact that the adult makes an effort to circumvent the "irregular" use of the verb by using the more marked (periphrastic causative) alternative (see Tomasello, 2011).

[9] The data obtained in this research can be observed through several different lenses: intra-constructional, inter-constructional, intra-age group and inter-age group observations. Inter-age group comparions are no doubt at the focus of this research, but they often cannot be interpreted in isolation from other construction and age-specific observations. In this sense, intra-constructional remarks delve into lexical and structural properties of the constructions observed primarily in relation to patterns occuring in them and the corresponding frequency trends, whereas inter-constructional observations concern differences across constructions, and not patterns occuring within them (see Footnote 1). Intra-age group data is a term used to capture all construction-specific results but within one age group, whereas inter-age group comparison takes all of the aforementioned into account to gain a better insight into developmental differences between the observed age groups in the language they produced.

[10] In linguistics, markedness as a concept has been referred to different phenomena in linguistics, but recently it tends to be used as the designation of cases which go against our expectations of what they should look like. A phenomenon is more marked than another if our general expectations regarding its possibility of occurring are

which the language acquisition process should be approached through the lens of frequency, and serves as an instruction for interpreting the results of this study:

> "And so our proposal is actually that frequency and maximal marking play very different roles in acquisition. Children will acquire first the instantiation of the construction they hear most frequently. But then they will bring together different instantiations of the construction on the basis of similar function, such that a prototype is formed including all of the marking options. And so we might actually propose two different routes children might use to get to their own prototype representation of a linguistic construction (as well as various possible mixtures of these strategies)."

Some approaches to language acquisition focus on the end result, while others emphasize the process of how children acquire language. Linguists tend to focus on the end result, such as the grammatical patterns, while psycholinguists are more interested in the ways that speakers access the pertinent structures and produce the utterance in a stepwise procedure by which the utterance is produced (Clark, 2009, pp. 10-12). In other words, psycholinguistic approaches to language acquisition tend to focus on the internal, cognitive processes behind using particular constructions such as the relative construction. These aspects of language acquisition remain beyond the scope of this study, although they are sometimes referred to as possible interpretation tools for the data retrieved. Instead, the focus of this study is on the actual language use in the context of language development and acquisition. Given that the emphasis is on the three constructions in terms of how they are realized, the approach undertaken in the thesis no doubt belongs more to the linguistic domain, though it is difficult to isolate the issues completely; the questions addressed often (at least to a degree) concern the pragmatics of language use, the need to use the observed patterns, the process of accessing and forming the constructions in question. Nevertheless, the linguistic approach is evident in the fact that the final product is in the focus of the study, i.e. the representation of the observed constructions in everyday speech with respect to their internal syntactic and semantic features. This is also the point where the thesis steps into an explorative domain where several different frameworks converge into one; that is, it mostly remains under the scope of Cognitive Linguistics, making

lesser, and thus closely correlates with frequency and distribution, but also conceptual and structural complexity (Gaeta, 2017).

use of the ideas such as the 'usage-based' and Construction Grammar (CxG)[11] approaches when it comes to interpreting language patterns and language production, but it also introduces corpus-based methodology as a tool to observe actual language use, and makes claim about the value of such findings. In solely observing the language produced, the corpus-based approach emphasizes the linguistic aspect of acquisition, while taking into consideration the social setting in which the language is produced, as well as the cognitive foundations that the observed age groups build upon (cf. Núñez, 2007). The comparison of early and adult language allows for conjectures about the processing of external linguistic input and input reproduction.

Aa a method, a corpus-based research entails the collection and analysis of large amounts of naturally occurring language data, or 'corpora', in order to study patterns of language use and meaning. One of the main advantages of corpus data is that it reflects how language is actually used in real-world contexts, rather than being based on artificially constructed examples or experimental tasks (see Section 1.3 for more details). On the other hand, the limitations are obvious in the sense that the researcher always has to bear in mind the difference between language competence and language production.[12] Still, the strongest argument in favour of corpus data is that it allows researchers to study a large and diverse sample of language, which is potentially useful for studying language development. Children's language development can be difficult to study experimentally, as it involves a wide range of factors and is influenced by a child's individual experiences and environment, whereas corpus data provides a rich source of naturally occurring language data that can be analysed to investigate how children's language develops over time, and how it is influenced by different factors. In the context of this study, the research places considerable emphasis on the comparative data, while emphasizing (dis)similarity and correlation solely in linguistic terms as children move towards adult-like utterances. Overall, the study also argues for the position that such approaches to actual

[11] Note that the term Construction Grammar tends to orthographically vary across different papers and studies, and though it tends to be written with initial capital letters, it can also sometimes take the uncapitalized form – especially when the author wants to emphasize the fact that it represents a family of approaches (sometimes dubbed "flavours") which share some core underlying principles, rather than a unified monolithic theoretical strain (cf. Goldberg & Suttle, 2010). It has also been referred to as a framework or a field in its own respect (for a recent example, see Ungerer & Hartmann, 2023).

[12] It needs to be noted that the claims on the usefulness of actual language use data, and the ability to generalise on it (no matter how restricted) is not necessarily a radical proposition. For example, Stefanowitsch (2020, p. 4) has argued that if we believe linguistic competence to be at least broadly reflected in linguistic performance, then it should be possible to model linguistic knowledge based on observations of language use. In this sense, corpus linguistics is in the same situation as any other empirical science with respect to deducing underlying principles from specific manifestations influenced by other factors.

language data can serve as a valuable tool for investigating language development because they provide a comprehensive and realistic view of language use, but also allow researchers to study a large and diverse sample of language data.

This study is based on the family of corpora called CHILDES (Child Language Data Exchange System), that is, on those studies which were conducted among children whose first language is English (see the full list of CHILDES corpora sources used under Section 8.2). CHILDES is currently defined as an organized database that contains written records of spoken language (see Section 2.3.2 for more) which were mostly contributed by researchers who collected the data for their own research purposes (cf. Corrigan, 2012; MacWhinney, 2015). Nonetheless, the type of research such as the one employed for the purposes of this thesis is not a complete novelty, with several studies having already utilized CHILDES corpora as a second-hand source of data to compare parental input and child language (see Liu et al., 2008; Goodman et al., 2008; Li & Fang, 2011; Shin, 2022). For example, a study by Li and Fang (2011), who utilized the Manchester corpus data (part of CHILDES), which consists of transcripts of English-speaking children aged between 1 year and 8 months to 2 years and 25 days, aimed to explore the relationship between the frequency of words in the language children hear from their mothers (input) and the words they produce themselves (output). Apart from having noted that children use more words with concrete and easily visualized meanings, the researchers also observed that the mere quantity of words spoken to children by their mothers was not significantly correlated with the number of words acquired by the children, which according to them, indicated that the comprehensibility of the input is crucial children's language acquisition; i.e. it is essential to make sure they understand what is being said to them, rather than just having them mimic sounds or words. A considerable overlap in caregiver input and child production when it comes to construction types expressing transitive events was observed by Shin (2022), while Goodman et al. (2008) found higher parental frequency to be associated with earlier acquisition within specific lexical categories, but also noted that frequency interacts with category, modality and developmental stage in its impact of the acquisition of vocabulary. Liu, Zhao and Li (2008) conducted corpus research that examined lexical composition patterns within the vocabularies of children and their caregivers in three different language groups (English, Mandarin, and Cantonese) in CHILDES family of corpora on age groups spanning from 13 to 60 months. They highlighted that language-specific differences in the linguistic input significantly influenced children's language output. This influence was reflected in the varying

percentages of nouns, verbs, and adjectives present in a child's productive lexicon at different developmental stages, indicating also that, as children continued to grow and develop, their lexical composition patterns gradually became more similar to those of their parents. Again, any evidence based on such corpora needs to be interpreted with care since it represents data collected from spoken records. There may be a considerable disparity between what speakers can produce, and what they do produce, given their language abilities. Nevertheless, the primary advantage of CHILDES is twofold: it contains transcriptions of natural conversations between children and parents, and the congregated corpora of numerous child language studies is compelling in its size. The English corpora alone comprise more than 22 million word tokens, which enables reliable generalisations to be made based on the data. The sheer size of such corpora, and the improbability of collecting reproducible and retestable findings from the accidental data, allow us to make reasonable inferences on linguistic competence and not just performance.

All in all, the primary aim of the study is to investigate the lexical and structural similarities among different age groups in terms of their preferred lexical and syntactic choices within observed constructions. This primarily includes determining the extent of similarity in verb and noun frequency across these constructions, examining whether there is a discernible difference in similarity between the 0-3 age group and the 4-6 age group, and specifically assessing if the 0-3 age group exhibits greater similarity to child-directed speech than the 4-6 age group. The thesis also intends to identify any differences in terms of animacy concerning key lexical items (nouns) used within the constructions. Measures of lexical richness and complexity are also included for the purposes of reflecting differences in language use across age groups. Finally, the data should also shed light on the variation in the degree of item-based character across said constructions.

The first chapter of the thesis serves as an introduction to the general framework of the research, providing key points about the theoretical background exploited in the study design and data interpretation (see incoming subsections of Chapter 1). Chapter 2 contains several important subsections that delve into the exploratory scope of the research, outlining the types of analyses to be conducted on the corpus data, the research questions and hypothesis, corpus description and the method by which the constructions were recovered from the transcriptions. The research results on the three constructions observed – ditransitives, periphrastic causatives, and relatives – are outlined in chapters 3, 4, and 5, respectively. Each of them deals with the

constructions separately, but in a rather systematic fashion; each construction-specific chapter can be further subdivided into sections that operationalize the construction observed, recapitulate the previous research on the topic, and lay out the results of the corpus analysis. For each of the constructions, there is an additional section which summarizes the main findings of the examination and reflects on the stated research questions and hypotheses. Chapter 6 moves away from intra-constructional data towards inter-constructional data and discusses the implications of these findings on the stated research questions and the overall understanding of this heterogenous and comprehensive dataset. Finally, it presents study limitations that are to be considered by the reader of this thesis, after which the conclusion (Chapter 7) once again responds to the stated research objectives.

1.2. Cognitive Linguistics

1.2.1. The framework of Cognitive Linguistics

Cognitive Linguistics is a field that emerged in the 1970s as response to the premises and inferences of both Generative Grammar and Montague Grammar (Evans, 2012, p. 130). The framework of Cognitive Linguistics started to form and evolve around the names such as George Lakoff, Charles Filmore, Leonard Talmy, Ronald Langacker and others (cf. Lakoff, 1982; Lakoff & Thompson, 1975; Filmore, 1975; Talmy, 1978; Langacker, 1987; etc.). Hence, Cognitive Linguistics emerged as not as a "single, closely articulated, theory", but rather as a "broad theoretical and methodological enterprise" (Evans, 2012, p. 130). Similarly to Evans, Geeraerts (2006, p. 2) also calls it a "flexible framework rather than a single theory of language", further stating that "it constitutes a cluster of many partially overlapping approaches rather than a single well-defined theory that identifies in an all-or-none fashion whether something belongs to Cognitive Linguistics or not." Perhaps his most apt description of Cognitive Linguistics is that of a "theoretical conglomerate", consisting of theories which diverge from traditional approaches primarily in their treatment of the interrelationship between grammar and meaning. Although this multiplicity and intermixture of various emerging approaches at the time might be confused with inconsistency, this is where, in fact, lies the strength of this theory. Despite the fact that a wide range of claims has been made within this framework, they are typically oriented towards different subsections of the language phenomenon, and they mostly complement one another. This thesis exploits the accounts that

position themselves well on the theoretical and methodological line between cognitive and corpus linguistics.

Sometimes, terminological confusions might arise in addressing of the terms cognitive and Cognitive Linguistics, and Geeraerts (2006) provide us with a clear distinction between the capitalized Cognitive Linguistics and the uncapitalized cognitive linguistics. According to him, uncapitalized cognitive linguistics marks any approach in linguistics which observes language as a cognitive phenomenon. In that sense, generative grammar also belongs to this broader category of cognitive linguistics because it "attributes a mental status to the language" as well, regardless of the fact that it views language as an autonomous system, described through formalized syntactic structures and rules. Cognitive Linguistics is taxonomically on the same level as generative grammar and represents only one approach within cognitive linguistics. Nevertheless, given that the term is now more frequently associated with the narrower definition of Cognitive Linguistics, and in the rest of the thesis, the references to the term 'cognitive' will be the ones relating to this 'theoretical conglomerate' that represents a move away from the previous approaches to grammar.[13]

Considering the multiplicity of ideas that exist with the cognitive framework, there is a need to establish what exactly these ideas have in common. According to Evans (2012), we need to start with two commitments: the *Cognitive Commitment* and the *Generalisation Commitment*. The former represents the idea that cognitive linguistics deals with descriptions and models of language by taking into account other cognitive and brain sciences. The latter is a commitment of linguists to describe the knowledge as a by-product or a result of general cognitive abilities. In other words, in order to earn the label 'cognitive', the theory must be in sync with what is known about the way in which the brain functions, and keep track of the new discoveries made by the fields of psychology, neuroscience and even studies on artificial intelligence.

As already stated, another important common trait of approaches within Cognitive Linguistics is the way in which meaning is approached. Language is primarily about meaning

[13] Here, "previous" primarily refers to the generative-transformational approaches to grammar. Although there are shared points of agreement between certain theoretical strains within the Cognitive framework (such as Construction Grammar) and generative-transformational approaches to grammar (both recognize the language as part of the cognitive system), the most robust points of departure are that the advocates of the latter view formal structures as independent from semantic and pragmatic aspects of discourse, and that they argue for a factual necessity of hard-wired knowledge of otherwise (according to them) unlearnable grammar (see Goldberg, 2006, pp. 4-5).

and the information conveyed by a particular language expression is at the core of linguistic analysis, Geeraerts (2006) outlines the treatment of meaning in Cognitive Linguistics through four key assertions: (1) linguistic meaning is perspectival, (2) linguistic meaning is dynamic and flexible, (3) linguistic meaning is encyclopaedic and non-autonomous, and (4) linguistic meaning is based on usage and experience.

That meaning is perspectival is evident in the fact that it both reflects and shapes the outside world. The way in which we express something is constrained by the conditions in which we find ourselves at the moment of speaking, and the interpretation of that expression must take context into account. That linguistic meaning is dynamic and flexible is evident in the way meaning tends to change and shape depending on the social or temporal context. The argument that it is encyclopaedic and non-autonomous is inspired by the idea that it cannot be isolated from the world in which we live, or from our cognitive capacities for that matter. It is encyclopaedic in that it depends on other knowledge that we have acquired thus far in the world. Finally, the idea that linguistic meaning is based on usage and experience is rooted in the way Cognitive Linguistics perceives the emergence of meaning – meaning of a particular expression or a linguistic item is attainable only through recurring experience of how it is used, in which context, for what purposes etc.

In the initial stages of these new theoretical assumptions, the term frequently encountered was that of Cognitive Grammar (CG), which somewhat constitutes a bedrock of Cognitive Linguistics. Nevertheless, given its uniformity and specificity, CG separates itself as only one model of grammar within Cognitive Linguistics (there is also Construction Grammar and Word Grammar; see Geeraerts & Cuyckens, 2007); in other words, CG represents a particular linguistic theory within Cognitive Linguistics, that can be regarded as radical due to the claims such as that of grammar being completely symbolic (i.e. filled with meaning) or that fundamental grammatical notions such as ‘verb’, ‘noun’, or even ‘subject’, “have unified conceptual characterizations” (Langacker, 2010). Although the emphasis is often on the term ‘cognitive’, notion which has become entrenched in linguistics as an epitome for the claims of language being an essential manifestation of cognition, CG designates a theory that bridges the gap between cognition and interaction. Langacker states that the relationship between cognition and language is evident in the fact that it emerges from more general phenomena, such as perception, attention and categorization, and consequently cannot be autonomous:

> "The concern with cognition—not shared by all strands of functionalism—is fully compatible with the latter's emphasis on social interaction. It is only through interaction in a sociocultural context that language and cognition are able to develop. By the same token, an account of linguistic interaction cannot ignore the assessment by each interlocutor of the knowledge, intentions, and mental state of the other, as well as their apprehension of the overall discourse context. In its basic principles, CG (despite its name) strikes what is arguably a proper balance between cognitive and interactive factors." (2010, p. 89)

Cognitive Grammar is thus only a small part of this theoretical conglomerate; Geeraerts (2006) states that there are, in fact, 12 major parts that create the basis for Cognitive Linguistics. Apart from Cognitive Grammar, there is also Construction Grammar, usage-based linguistics, schematic network, prototype theory, conceptual metaphor, grammatical construal, radial network, image schema, metonymy, mental spaces and frame semantics.

In the context of this thesis, the part related to theoretical background, operationalization and interpretation, mostly draws from studies in Construction Grammar and usage-based linguistics. Let us then briefly define what these two terms relate to and why they are relevant for this research.

1.2.2. Construction Grammar

Much like Cognitive Grammar, the term Construction Grammar (CxG) is generally used as an umbrella term for a number of theoretical strains that share some key underlying assumptions pertaining to the lexico-syntactic continuum. One of the most prominent is the Goldbergian one, later known as Cognitive Construction Grammar (CCxG), which only represents one 'type' (see Boas, 2013, 2021). Other prominent accounts include Unification Construction Grammar (UCxG), Cognitive Grammar (CG) and Radical Construction Grammar (RCxG) (for an overview, see Goldberg, 2006, pp. 213-214). As can be assumed from the name, the priority is given to the term 'construction', or its more elaborate counterpart 'grammatical construction'. Hence, the first radical idea of Construction Grammar, that distinguishes it from its competing theoretical frames, is a move in the perception of grammar, and its interrelationship with semantics. First, the construction as a term that represents a particular syntactic pattern, started to be treated as a primary unit of a language. The crucial notion here

is that of a ‘unit’, because it immediately hints towards the existence of something that is inherent solely to the ‘unit’ in question, i.e. there is something that can be ascribed to the construction as a whole. If, in fact, we start to treat a construction as an entity itself, we must immediately recognize that there are characteristics that cannot be deduced from its components, and thus the contribution of the construction to meaning (or something else) can be detected on a greater level than on that of its constituents. In other words, this is where Construction Grammar differs from generative so-called componential models, where words are typically interpreted as associations between meaning, syntactic category and form (Croft, 2007).

In the early efforts to capture the distinctions between the emerging Construction Grammar and its rival theories, Charles Filmore (1988) laid out several arguments which he found in common with other syntacticians working within CxG:

> (1) Construction Grammar does not have transformations like transformational grammars, i.e. relationships that used to be captured by derivations of sentences and their structure are now captured by grammar as a whole;
>
> (2) Construction Grammar differs from simple-phrase structure grammar in that its structural categories may store complex bundles of information, and perhaps most importantly, it allows for a particular linguistic expression to be able to simultaneously instantiate more than one grammatical construction at the same level;
>
> (3) Construction Grammar differs from the generativist tradition in that it attempts to simultaneously describe syntactic patterns and the semantic and pragmatic purposes to which these patterns are dedicated.

Probably, the most important distinguishing feature of Construction Grammar is its treatment of syntax and lexicon. Similarly to what has already been stated about the interrelationship between grammar and meaning within the framework of Cognitive Linguistics, lexicon is not treated separately from syntax. While in the generative tradition, lexicon is viewed as a component in itself, in the Construction Grammar it is viewed as a part of the continuum which encompasses both lexicon and syntax. The inevitable conclusion is then that the only difference between lexicon and constructions is in degree (Croft, 2007, p. 470-471). The difference between words and constructions is in syntactic complexity; constructions may be made up of words or phrases, while words are syntactically simple. According to Croft (ibid.), if we take an example of morphologically complex words, the interpretation of Construction Grammar

would be to treat them as constructions in themselves, the only difference being is that its parts are morphologically, not syntactically, bound. Finally, Croft (2007, p. 471) summarizes the most important point of Construction Grammar in the following claim:

> "The end point of this argument is one of the fundamental hypotheses of construction grammar: there is a uniform representation of all grammatical knowledge in the speaker's mind in the form of generalized constructions. The constructional tail has come to wag the syntactic dog: everything from words to the most general syntactic and semantic rules can be represented as constructions. Construction grammar has generalized the notion of a construction to apply to any grammatical structure, including both its form and its meaning."

Again, much like Cognitive Linguistics, Construction Grammar is not immune to the idea of cognition being inseparable from the ways in which language operates and how it is processed. When talking about the interrelationship between meaning and form, Construction Grammar in fact addresses the way in which units are represented in the mind of the speaker as well. Under the assumption that the interpretation of such interrelationships is related to language processing and the ways in which we construct mental representations of particular expressions, the question that becomes central in the context of language acquisition is how these expressions are stored, processed and learned by children. Can a particular construction be stored much like a word would be stored in the mental lexicon? How does the level of schematicity these constructions allow change in the context of development? Can children recognize the meaning that is inherent solely to construction, as something separate from the sum of its constituents? Are these constructions learned imitatively and how do these processes change with the progression of age?

1.2.3. Usage-based linguistics

The second prominent set of ideas that has evolved into a theoretical framework in its own right is usage-based linguistics, a highly useful and appropriate frame of reference for this thesis given that it also situates itself within the wider Cognitive framework. In usage-based linguistics, the propositions of interrelatedness between cognition and language have been extended to the ideas of interrelatedness between language use (performance) and language competence, a radical idea that separates it from the structuralist and generative tradition. It is

not just that language performance is a product of language competence, but it is also the other-way around. In fact, the emphasis of usage-based linguistics is that the speakers' knowledge of language, the way it is organized, structured or mentally represented, is the product of language use and performance (Diessel, 2017). Within this framework, speaker's or learner's experience occupy a central position. In return, this experience influences the learner's mental representations of language.

Apart from the emphasis on the fact that communicative events drive language acquisition, influence its structure and organization, the usage-based theory starts out with the claim that "language is an extension of other cognitive domains" (Bybee & Beckner, 2010, p. 829). This implies that language needs not be considered as a separate "module" (cf. Fodor, 1983), or that the capacity for language has its own mechanisms, unrelated to those of other cognitive domains. Again, this is another claim that establishes this approach within the cognitive framework, but more importantly, this part enables the linguists working within the usage-based linguistics to explain how exactly the reproduction of the input occurs. Cognitive abilities such as statistical learning and sequential learning, chunking and categorization, are in fact general capacities of the human brain from which the mechanisms of language are derived (Bybee & Beckner, 2010). In other words, these capacities are at the core of language processing and acquisition.

An important part of the usage-based approach is also that it does not take a reductive approach to language memory and storage. It is assumed that people can retain a rich memory of representations, and the language input they receive during their lifetime is stored in the form of exemplars (Bybee & Beckner, 2010, p. 833). An exemplar represents a word or a phrase heard in a particular instance and remembered by the learner. According to this model, there is a cluster of exemplars that learners remember, and based on the frequencies of particular exemplars, their knowledge is either reinforced or expanded with new exemplars that position themselves closely to other similar exemplars in a metaphoric mental cluster of these instances. This also means that the mental representation of a particular word, phrase, or even construction will be influenced by every token of use (Bybee & Beckner, 2010, p. 833). When speakers produce their language, they choose from one of the exemplars previously encountered in the input. This is not to say that language productivity is to be completely ignored at some point, but the emphasis is on the reproductive character, especially when it comes to early stages of language acquisition.

The balance between the imitative part of the acquisition process and the role cognitive abilities in the usage-based model of language is perhaps best explained by Tomasello (2001a), who employs it to put forward 5 separate observations regarding language acquisition:

> (1) child language acquisition relies on the utterance as the primary psycholinguistic unit, which implies understanding communicative intentions conveyed by the expression at the foundation of the utterance;
>
> (2) in the initial stages of language acquisition, children strive to reproduce entire utterances, rather than just single words;
>
> (3) the early utterances produced by children are reproductions of the instances encountered in the input, i.e. their language can be explained in terms of concrete 'item-based' schemas and constructions;
>
> (4) children observe the type variations in parts (or 'slots') of the heard expressions/constructions and are able to generalise and draw abstractions, i.e. children are able to make "functionally based distributional analysis" and based on it, create novel utterances;
>
> (5) children first rely on utterance-level schemas in their syntactic operations, but they create novel utterances when they modify those schemas according to the needs of a particular communicative situation.

Imitation is understood to play an important part in language acquisition according to this model, but not in the sense that children's language is parrot-like where the child just reproduces something they heard. The important addition to the theory is pairing this imitative part with the cognitive aspect. When the children hear the utterance, their intention is presumably different from that of a parrot – they strive to understand the utterance instead of just reproduce it. This is where the memory start to play a part and the early imitative utterance becomes simply a by-product of insufficient experience. In order for the child to be productive, the assumption is that it first needs to hear a sufficient number of expressions in order to become aware that certain 'slots' can be filled with various lexemes. Experience and cognitive abilities help bridge the gap between imitation and creativity/productivity. The key points of the usage-based approach to language acquisition are perhaps best summarised by Tomasello:

> "In the early stages, children mostly use language the way they have heard adults using it. This leads to an inventory of item-based utterance schemas, with perhaps some slots in them built up through observed type variation in that utterance position. The reason

> that children do not operate with more abstract linguistic categories and schemas is quite simply because they have not yet had sufficient linguistic experience in particular usage events to construct these adult-like linguistic abstractions." (Tomasello, 2001a, p. 70)

And also:

> "Fluency with a construction is a function of its token frequency in the child's experience (entrenchment); creativity with a construction emanates from the child's experience of type variation in one or more of its constituents (abstraction). In this way, children build up in their linguistic inventories a very diverse set of constructions – concrete, abstract, and mixed – to call upon as needed in particular usage events." (Tomasello, 2001a, p. 76-77)

In other words, from the investigative point of view, there are two ideas that can be exploited in this corpus-based study. Experience translates to the frequency of occurrence, because in the input children receive, it is the frequency that solidifies the representation of particular words and expressions in the memory of a child, and it is the frequency that facilitates the processing of these linguistic elements (Diessel, 2007). Frequencies comprise a significant part of the data presented in this thesis, with the goal being to illuminate inter- and intra-group differences in the production rates of specific structures and lexical items, and interpret the findings in light of early language development. In this light, the fact that usage-based approaches treat syntactic structures as lexically particular is particularly important (Diessel, 2007). Early reproductions of utterances follow these item-based patterns, which means that the expressions reproduced early on ought to mirror the same lexemes encountered most frequently in the everyday input with respect to the constructions in which they occur.

1.3. Corpus linguistics

1.3.1. Corpus-based and corpus-driven methods

Despite the fact that it has been present in linguistics for quite some time, the term corpus linguistics does not have a clear-cut definition. The ambiguity stems from the fact that some perceive it as a methodological choice, while other view the term as a label for a theory (Gilquin, 2010, pp. 5-23). The question is whether doing research on corpora makes someone a corpus linguist, whether the research is only backed up by the corpus data or whether the

theoretical framework is dependent on the methodological one to the point that it constitutes a theory of its own.

Probably the main distinction methodology-wise in the field of corpus linguistics is the approach linguists tend to take towards the data itself, i.e. the divide is typically between the emphasis on quantitative data vs. the qualitative data. Thanks to advancements in data storage and language processing tools, it has become easier to capture the valuable quantitative data that was inaccessible before. Nowadays, the research done in corpus linguistics is characterized by the machine-readable corpora, and the technological advancements in the field have granted the linguists the opportunity to use the reliable part-of-speech tagging (POS tagging). Nonetheless, Gilquin (2010, p. 8) warns that it is sometimes essential to go beyond numbers to gain deeper theoretical understanding:

> "Frequencies and statistics are interesting, and certainly represent one of the greatest insights into language that corpus linguistics has made possible. However, they should not be seen as an answer, but rather as an incentive to ask questions. The quantitative perspective, in other words, is not an end in itself, but a starting point for qualitative research […] It is therefore crucial that bottom-up research does not confine itself to number crunching, but goes all the way from the data to theoretical considerations."

The research in corpus linguistics can be divided on similar grounds as in any other type of research: deductive and inductive. Although the terms that correspond to inductive and deductive methods in corpus linguistics have been previously used, the distinction between the so-called corpus-based and corpus-driven research has been elaborated by Tognini-Bonelli (2001). Thus, inductive research in corpus linguistics is called corpus-driven research, and the main contribution of this type of research is to identify the existence of various linguistic patterns or constructs in language that have eluded some earlier linguistic theories (Biber, 2010, p. 160). On the other hand, corpus-based research is a term that corresponds to deductive type of research. Biber (2010, p. 162) claims that "corpus-based research assumes the validity of linguistic forms and structures derived from linguistic theory; the primary goal of research is to analyse the systematic patterns of variation and use for those pre-defined linguistic features." The difference then is that corpus-driven analysis starts out with minimal assumptions regarding the investigated linguistic features, relying more on the patterns emerging from analysis as co-occurrences of particular words. Therefore, the starting point of the analysis used in this thesis corresponds more to the corpus-based research, considering that the investigated

phenomena is first defined on the basis of different syntactic and semantic parameters, and only then targeted and located within the corpora.

When it comes to the investigation of different language varieties, it has been frequently claimed that the limitations of corpus-based analysis are that the findings of particular language variation and use are not necessarily valid for a language as a whole. The problem arises from the concern that has been emphasized more in the context of corpus-based research – the representativeness of the corpus itself. This is especially relevant in the context of language acquisition and the corpus-based exploration of the same.

According to Biber (2010, p. 163), the evident benefit of corpus-based research has been its reveal of the dramatic differences found across different varieties when it comes to particular linguistic features and variants. The results of such research often confronted the universal descriptions of language, thus emphasizing the importance and value in studying varieties and taking into account distinct linguistic variants in the development of linguistic theories. Corpus-based research is basically designed to test the predictions that linguists might have on the language use – and more often than not, the predictions based on intuitions can turn out to be wrong. For example, the study by Biber et al. (2000, pp. 460-463) showed that despite the assumptions from linguists that a progressive aspect is the most common one in English conversation, it turned out that the simple aspect was most frequently used by far. Furthermore, Biber observes that, despite expectations that students use dependent clauses more in their writing than in speech, and that this tendency progresses as they develop their writing skills (see Wolfe-Quintero et al., 1998), corpus-based studies have shown that these predictions about their language use were, in fact, wrong (Biber, 2010, p. 166). The use of dependent clauses can be interpreted as a measure of grammatical complexity and the natural expectation is that students ought to demonstrate greater grammatical complexity in writing than in speech. While the use of relative clauses was greater in academic writing, other dependent clauses such as finite adverbial clauses and complement clauses were used more frequently in conversation.

The present investigation was conceived as a corpus-based study aimed at elucidating certain established hypotheses in language acquisition theory. Although the results found in this study are extremely valuable for language acquisition researchers, some of them give rise to further inquiries rather than unequivocally resolve existing questions. Here, the corpus inquiry is primarily designed to challenge assumptions regarding early language use, as well as to

investigate (dis)similarities between adult and child language use, whilst laying groundwork for future investigations into the phenomenon of language acquisition.

1.3.2. Corpus linguistics and Cognitive Linguistics

Given that this thesis mainly relies on the framework of Cognitive Linguistics, it appears useful to explain how this framework fits alongside corpus linguistics. First, as already noted, this study methodologically relies on corpus linguistics, i.e. although the term has been used to represent more than just mere methodology, this research builds upon theories surrounding first language acquisition and treats corpus approaches primarily as a means to an end.

Although nowadays, the fact that corpora play a significant part in the research of linguists working within the cognitive framework might seem ordinary, this was not necessarily always the case. Cognitive Linguistics started to emerge around 1980s, and although it differed significantly in terms of its theoretical foundations and postulates from the traditional generative accounts, it still continued with the similar methodological restraints towards the linguistic data that is to be included in the research (Stefanowitsch 2011). It is often said that Cognitive Linguistics elaborates on the linguistic categories before engaging the data, that is, it defines categories beforehand, while corpus-driven linguistics induces categories from the data itself (Teubert, 2008). At first glance, these two seem to be the opposites that might not unify together well, but it is precisely the combination of the two that gives credit to both. The absence of data is complemented by a comprehensive theoretical overview in one, while the absence of theory is complemented by the abundance of data in the other. In her explanation of why these two work well together, Gilquin (2010, p. 15) claims:

> "It is true that corpus linguistics extracts categories from the data and cognitive linguistics posits them beforehand, but this is no different from saying that corpus linguistics starts with the data, while cognitive linguistics starts with the theory. What is crucial is that both paradigms recognise the existence of categories. More fundamentally, they both see category membership as a matter of degree rather than a yes-or-no question. In cognitive linguistics, this principle is expressed by means of the concept of prototypicality, originally a concept from psychology, according to which categories are organised around a maximally representative example (the prototype). Depending on the similarity they exhibit with the prototype, the other members occupy

> a more or less central position within the category. As a consequence, categories have no clear boundaries and overlap with one another. It is the same fuzziness that corpus linguistics has revealed."

The idea of unifying cognitive framework and corpus studies has probably been best represented by the 'usage-based' models in modern linguistics (see Barlow & Kemmer, 2000; Mukherjee, 2004; Tummers et al., 2005; Gries & Stefanowitsch, 2006; Heylen et al., 2008 etc.). In this context, the term 'usage-based' seems to be the common denominator between the two theoretical and methodological frameworks. In fact, in their definition of Cognitive Linguistics some argue that Cognitive Linguistics is usage-based by default; for instance, Glynn (2010, p. 89) states that:

> "Cognitive Linguistics is, by definition, a usage-based approach to language. Its model of language places usage at the very foundations of linguistic structure with a linguistic sign, the form-meaning pair, argued to become entrenched through repeated successful use. It is this entrenchment that renders symbolic gestures linguistic rather than merely incidental and represents the key to structure in language. Patterns of language usage across many individuals can be argued to be indices of shared entrenchment. When large numbers of language users possess the same or similar entrenchment, we can talk about grammar, that is, linguistic structure."

Corpora thus serves to attest the use of language in various settings, contexts and manners, relying on the shared properties between language representation and language manifestation. The usage-based model in many ways represents a bridge between the tools and methodology of corpus linguistics and the theoretical account of Cognitive Linguistics.

1.3.3. Extracting the data – corpus as a method

Considering that the three constructions targeted by this research are structurally and lexically different, different methods were used for their extraction from the corpora. The extraction of ditransitives, periphrastic causatives and relative constructions, was conducted performing queries which exploited either the exact word anchors or POS anchors. Given the flexibility and openness of ditransitive construction to accept new verbs (Goldberg, 1989, 1992), the construction was extracted solely through the use of its syntactic pattern. For periphrastic causatives and relative constructions, the extraction relied on both specific lexical

items and part of speech category. The latter approach is less inductively prolific (especially when it comes to the periphrastic causative), since it does not reveal new lexemes as heads of these constructions. Nonetheless, the mixture of the syntactic pattern along with the particular lexemes in the extraction process, although not ideal, helps to reduce the redundant patterns that make the filtering operation considerably more difficult and time consuming. This leads us back to the differences between the use of corpus-driven and corpus-based research, and detecting where this thesis positions itself on this continuum. In terms of data extraction from the corpora, Biber (2010, p. 162) states that "corpus-driven analysis assumes only the existence of word forms". In other words, such research would rely on targeting particular lexical bundles or word sequences (for instance, one may try to discover the frequency of the sequence *never have I ever* in a particular corpus). Corpus-driven approaches are thus primarily defined by their focus on particular word sequences, that are rarely structurally complete or idiomatic, and on the analysis that is based solely on the rates of particular occurrences and their distribution across different texts. In terms of just data extraction, it would appear that this research does not start out as a corpus-driven research, but rather belongs to a more hybrid type of form that is also referred to as 'pattern grammar studies':

> "The pattern grammar studies might actually be considered hybrids, combining corpus-based and corpus-driven methodologies. They are corpus-based in that they assume the existence (and definition) of basic part-of-speech categories and some syntactic constructions, but they are corpus-driven in that they focus primarily on the construct of the grammatical *pattern*…" (Biber, 2010, p. 175)

Basically, these studies are corpus-driven because they focus on the linguistic units that emerge from corpus analysis, but start from the position of corpus-based studies as they take into consideration the pre-defined constructions that lead the extraction process. This hybrid approach to corpus is also that one that bridges the gap between corpus linguistics and constructional approach in the cognitive framework. The research must, at times, employ particular lexical items because relying solely on syntax would be impractical to the point of inadequacy, but there also needs to be room for syntax and its contribution to the meaning of the construction, particularly from the constructional perspective. This is perhaps best illustrated by the quote of Hunston and Francis (2000, p. 3):

> "Patterns and lexis are mutually dependent, in that each pattern occurs with a restricted set of lexical items, and each lexical item occurs with a restricted set of patterns. In

addition, patterns are closely associated with meaning, firstly because in many cases different senses of words are distinguished by their typical occurrence in different patterns; and secondly because words which share a given pattern tend also to share an aspect of meaning."

This methodological position is inspired by the idea that particular grammatical patterns are tied to particular meanings. For example, the ditransitive pattern is that of a *subject – verb – object – object*, and by targeting the word classes that occupy such positions (for instance, *PRO – V – NP – NP*) we ought to retrieve the instance with particular verbs that do not usually carry ditransitive meaning. The role of such research becomes partly to confirm that syntax can dictate the meaning, and not necessarily just the lexical items. While the concerns around syntax-semantics interface are not the primary focus of this investigation, it remains to be noted that such approach to corpora is at the methodological core of this thesis.

1.4. Three constructions – the increase in complexity

As already stated, in order to tackle the acquisition of the ditransitive, causative and relative patterns, the thesis approaches the matter by using evidence from corpora whilst interpreting the extracted data with the help of cognitive theoretical framework, meaning that the aforementioned patterns will thus be referred to as constructions (cf. Section 1.2.2).

As already discussed in the introduction, the move away from the traditional perception of meaning being restricted to lexicon started to change with the first observations that syntax can influence meaning. Although the idea was popularized by Goldberg (1995), there were many before her arguing for the interrelatedness between syntax and meaning (Givon, 1985; Langacker, 1985; Clark, 1987; MacWhinney, 1989 etc.). Even back in 1968, Bolinger observed that any change in the syntactic form must imply a change in the meaning (p. 127), which ultimately drove Goldberg to reflect on their findings and postulate a "Principle of no synonymy" (1995, p. 67): "If two constructions are syntactically distinct, they must be semantically or pragmatically distinct". Because of this, the constructions targeted by this thesis will be analysed primarily based on their different syntactic patterns. In a certain way, this principle leads us to assume that even minute differences in syntax can in fact influence the interpretation of these constructions, i.e. different patterns with slightly altered syntax that appear within ditransitive, causative or relative constructions can be regarded as mini-

constructions in themselves. Although the starting point of this work, these assumptions are not the object of this study in themselves, nor are the taxonomic differences between the patterns that we find among the family of constructions targeted in this research. Given that this thesis does not deal directly with detailed theoretical and syntactic accounts of a singular construction, or solely with its interpretation, it needs to be noted that the emphasis will be on the extracted data and the possible conclusions that might be drawn from it, and not on the presuppositions of any theoretical account about a particular construction and what its variants might entail. Nevertheless, the fact that certain syntactic patterns are more frequent than others within certain age groups reveals something about the processing difficulty of the same.

In the research, the term 'construction' is used for all three grammatical patterns observed in this study. While the term is more frequently used in the Goldbergian sense with ditransitives and periphrastic causatives, especially when it comes to research in Construction Grammar, relatives tend to be referred to as clauses (relative clauses, however, only constitute a part of the relative construction). Before proceeding, it is necessary to define what exactly constitutes a construction in this particular context. As already stated before, even words are conceived as constructions from the point of view of Construction Grammar (see Croft, 2007, p. 471). Any syntactic pattern which has adopted a "conventional function in a language" can be regarded as a grammatical construction, i.e. the fact that its structure is associated with a particular meaning and that this association has become linguistically conventionalized makes it a construction (Filmore 1988). This 'construction' can be described through terms such as external and internal syntax. The former signifies speakers' knowledge about the properties of a particular construction, and how they might be accommodated by a wider syntactic contexts. The latter term (internal syntax) signifies the construction's make-up, that is, the description of the construction in terms of its grammatical pattern (ibid.). Considering that constructions are viewed primarily as "form-meaning correspondences" (Goldberg, 1995), the label construction is more often reserved for the syntactic patterns which obviously contribute to the meaning of a particular sentence, that is, even when the participants of the particular construction are replaced by non existing words (or even verbs for that matter), there is still a meaning that can be expected from that particular construction. In other words, the non-compositionality principle suggests that the construction can be regarded as a construction if its form or meaning cannot be predicted solely by the component parts which constitute it (Goldberg, 1995, p. 13). This phenomenon has particularly been investigated and corroborated on the ditransitive

construction (Goldberg, 1989, 1992), while the periphrastic causative construction has also been similarly approached (Gilquin, 2006, 2017).

One of the reasons why these are often referred to as constructions is partly because one can imagine that a learner might store these as units, considering their degree of prototypicality and structural coherence, whereas relative constructions are on a different syntactic level from ditransitives and are often referred to as complex sentences (see Diessel, 2004). On the other hand, if we take a sentence such as *The man gave the boy a ball which was black*, it is hard to determine which construction is superior to another in terms of syntactic complexity. If anything, it appears that it is the ditransitive construction which accommodates the relative clause, and in the same sense accommodates a relative construction (*a ball which was black*). This happens because relative clauses act similarly to modifiers in a sentence, allowing themselves to be attached to any noun phrase. In terms of Cognitive Linguistics, relative clauses exhibit high schematicity in that they can be accommodated by various schemas, and yet their presence transforms the entire construction into a relative one. Similarly, if we look at the sentence *The man forced the boy to throw the ball which was black*, the question that arises is whether we should interpret this sentence primarily through the lens of a periphrastic causative or of relative construction. This however, does not represent a problem for Construction Grammar, which, in allowing polysemy within a family of constructions, must also allow for constructional polirepresentations at the same time. In this sense, it is also important to distinguish relative clauses from relative constructions. While the relative clause is an integral part of the relative construction, it is only in its entirety that we can argue for an advancement in syntactic complexity when compared to ditransitive or periphrastic causative constructions. Although this thesis addresses the suppositions concerning gradual complexification of the children's grammar, the benefit of the study will be the additional evidence pertaining to the differences in complexity between the observed constructions.

The term 'complexity' itself has proven problematic to clearly define, with various interpretations and uses by scholars. For example, 'linguistic complexity' has been used in different contexts for different purposes, some of which saw the construct as being intricately linked with sentence processing (i.e. the degree of computational resources a structure requires) (see Gibson, 1998), while others used it as a "cover term" for a variety of factors that include syntactic, thematic and semantic complexity (see Friederici et al., 2006). More straightforward operationalisations of the term moved away from associations with difficulty, and instead saw

the amount of information needed to describe it as an objective property of the structure in question (Dahl, 2004, p. 2). Building upon that, Odlin (2012) used a similar approach by referring to it as "descriptive complexity", and later concluded that "linguistic complexity" is best defined as the level of descriptive detail required in accounting for all phenomena that may present a challenge to learning. On the other hand, Trudgill (2017) used the term in the context of sociolinguistic typology where he defined it as a multifaceted concept, arguing that complexification is best defined as an opposite process of simplification, i.e. opposite of regularization, reduction of redundancy, and an increase in transparency. Factors contributing to linguistic complexity thus include greater irregularity, heightened syntagmatic redundancy such as repetition of information, increased use of morphological categories, elevated levels of allomorphy, and a higher degree of fusion.

When it comes to the analysis of intra-constructional differences and the specific syntactic patterns within each of them, the use of the term in this thesis is closest to the one presented by Pallotti (2015), who defined it purely in structural terms, where complexity stems directly from the quantity of linguistic elements and their interconnections (see also Arnold et al., 2000, where the term has been operationalized based on the length of constituents). Such operationalization deliberately excludes considerations of cognitive cost (difficulty) and developmental dynamics (acquisition), and instead takes into account the length of phrases, the number of phrases within a clause, the quantity of clauses per unit, the number of word-order patterns, and the mean length of clauses.[14] For instance, the use of the term 'complexity' in ditransitive constructions is mainly linked to the nature and length of the construction components, such as nominal elements,[15] and it is referred to as 'structural complexity'. In this context, the intricacy of individual elements is closely connected to the overall complexity of the construction, and the same logic is applied for other constructions. The reference to the complexity of the entire construction builds upon the complexity of its nominal elements, and it is not to be conflated with processing difficulty. For example, in relatives, the studies of relative constructions tend to concentrate on that particular aspect (see Section 5.1.3), i.e.

[14] It is essential to acknowledge, however, that even seemingly simple approaches like word count-based definitions can pose challenges, as exemplified in cases where ambiguity arises regarding whether a word should be counted as one or two, such as in the instances of "sports car" or "sportscar" (Pallotti, 2015, p. 123). In English, the somewhat common discrepancy between phonological and orthographic complexity/length also needs to be recognized as a factor.

[15] Note that research has demonstrated how the length of arguments influences the selection of dative realization (see Bresnan et al., 2007).

employing a pronoun in a construction instead of a modified noun does not necessarily make it less complex in terms of processing, as seen in the contrast between the cognitive load required to process the relativized noun *the car which was driven* and the relativized pronoun *it which was driven*. On the other hand, the thesis also makes use of term 'lexical complexity' as described by Lu (2012), which is further elaborated in Section 2.1.3. Though processing difficulty remains beyond the scope of this study, and though the term is primarily used to denote an increase in linguistic material when it comes to specific intra-constructional comparisons of pattern frequencies, the results provide a solid foundation to further one's inspection into the reasons for certain trends in production of specific patterns and the related implications on developmental dynamics and cognitive cost in general.

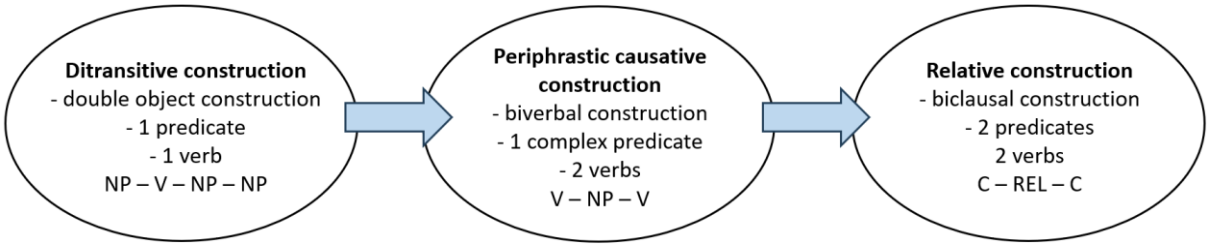
As far as inter-constructional differences and the use of the term complexity are concerned, the fact that there may be more to it than simply an increase in (or complexification of) linguistic material calls for consideration. For instance, the periphrastic causative construction is assumed to be more complex than the ditransitive construction because the predicate is more complex. On the level of meaning, the periphrastic causative also implies a greater number of agents in the event; given that the action is governed by two entities in the process, we may even assume the entailment of two events on the metaphoric level of the construction. Whether with a full infinitive or a bare one, the periphrastic causative is a biverbal construction and it entails two events – one which precedes and causes the other to start. The syntactic complexification is followed by the complexification in meaning, and the comparison might remind of Diessel's claims about the complexification of the nonfinite clauses in early children's speech. Diessel (2004) distinguishes between children's early nonfinite complement clauses based on the roles of the subject and the direct object in these constructions. He concludes that the infinitival extension of the nonfinite clause can be interpreted as a complexification of the initial expression, and that in the process of acquisition, children tend to learn this incrementally.

> "Moreover, while children's early nonfinite complement clauses are always controlled by the matrix clause subject (e.g. ***I*** *wanna sing*), later nonfinite complements are often controlled by the direct object, which occurs between the finite verb and the infinitive or participle (e.g. *He told* ***him*** *to leave*). Based on these findings, I argue that the development of nonfinite complement constructions can be seen as a process of clause expansion. Starting from structures that denote a single situation, children gradually

> learn the use of complex sentences in which a nonfinite complement clause and a matrix clause express a specific relationship between two states of affairs." (Diessel, 2004, p. 49)

Diessel's inference about the increase in complexity is applied to constructions where one clearly behaves as an extension to the other. This is evident in their interpretation, the increase in syntactic complexity in terms of argument and predicate structure and the fact that one can be a direct extension of the other (*He told him to leave* instead of just saying *He told him*). Although there is no direct implication of the extension, the argument of the increase in complexity is similar to Diessel's. This is not to say that there can be no overlaps between the ditransitive construction and the periphrastic causative. If we take Diessel's example *He told him to leave*, we can argue that the construction shares features with both ditransitives and periphrastic causatives. Similarly to ditransitives, there is something (a message) that is being transferred between the agent and the recipient. Similarly to the periphrastic causative, there is an implication (although causation is not guaranteed) that the occurrence of one event leads to the occurrence of another ('his suggestion leads to a departure'). Clearly, there are points of overlap where it becomes difficult to assess the difference in complexity, but it is still justifiable to argue that the aforementioned ditransitive construction (if we grant it such a label) is more complex than the prototypical construction with *give* (as in *He gave him the ball*). The point is that these discrepancies can be captured by the corpus-based approach, and that they may reveal interesting new data about language acquisition and processing. Figure 1 illustrates the operationalization of the increase in complexity in the three constructions targeted by this research.

Figure 1. Graphical representation of the increase in complexity across the three constructions



Legend: **NP** – noun phrase, **V** – verb, **C** – clause, **REL** – relativizer

Thus far, the research in language acquisition has primarily been focused on the differences in the acquisition difficulty between similar patterns within a family of constructions; for instance, with different types of ditransitives (see Gropen et al., 1989; Campbell & Tomasello, 2001; Snyder & Stromswold, 1997¸ Fischer, 1972 etc.) or relatives (see Diessel & Tomasello, 2005; Diessel, 2009a). Such studies tend to observe differences in the speed of mapping, the relevance of the input, the acquisitional time frames for these patterns etc. Although this thesis mainly investigates the range of frequencies between different syntactic patterns within each of the construction families observed, the chapter following the intra-constructional (i.e. construction-specific) analysis evaluates the discrepancies between all three observed constructions (for inter-constructional data, see Chapter 6).

2. RESEARCH PURVIEW AND METHODOLOGY

The following sections outline the aspects of language use observed across age groups, the specific measures employed in the research, the research questions addressed by the data in question, as well as the method by which it was retrieved.

2.1. Exploratory scope of the research

2.1.1. Structural similarity and complexity

A significant portion of the data provided in this study deals with the correspondence in frequency between different forms of the three constructions studied. The observation of similarities in the overall production of different patterns in the context of input relevance, and especially in connection to the lexical items found in the same forms, expands on the research that has explored the connection between form, function and the role of frequency (see Lieven, 2010 for a detailed overview). For example, an important insight on the topic was shared by Cameron-Faulkner et al. (2007) who found that the frequency of a particular form in the input that expresses a single function will facilitate the acquisition of that particular form (it is produced earlier by children), as opposed to when the single function is expressed across a number of different forms (in which case the child would use more generic, multi-purpose expressions. Although the data in this thesis is not retrieved for the purposes of addressing questions that would otherwise require longitudinal design, nor is it conclusive in regard to which forms arise in speech before others, the correspondence between the frequency of very specific forms and very specific items within those forms additionally sheds light on the role of the input and on the children's tendency to (re)produce the language material at hand.

Much of the research on constructions has already intersected with language acquisition studies and each of the constructions observed in this research has already been tackled with regards to some of its lexical and structural characteristics in light of language development (some of the relevant previous research on ditransitives, periphrastic causatives, and relatives, is outlined in sections 3.1, 4.1, and 5.1, respectively). When referring to the structural characteristics of the targeted constructions, the text addresses primarily the syntactic aspects of each of the constructions, their variations in article/noun/pronoun uses across age groups,

but it also includes some of their specific head lexical items (such as periphrastic causative verbs or the relativizers in the relative constructions). Although this will be noted several times throughout the thesis, it is important to note that each separate case study of the constructions is limited in some sense; for example, the observation of ditransitives covers only non-prepositional realizations, the observation of causatives only periphrastic realizations and the observation of relatives only constructions with overt relativizers. As much as it relies on previous theoretical considerations, the observation of their structural aspects is equally defined by the methodological boundaries of this work, i.e. much of what has been addressed concerning the structural aspects of the constructions has also been limited by what is possible (or achievable in a realistic time frame) to extract from the corpora by using POS tagging options. Correspondingly, the comparison of the structural aspects of the constructions exploits their variations in noun phrase architecture (such as their complexity in ditransitives), their verb phrase architecture (such as infinitival alternations in periphrastic causatives), or the overall alternating patterns of key syntactic elements (such as the difference between subject and object relative constructions). The discrepancies in the use ratio of particular forms across age groups would indicate the degree to which other factors (beside the input frequency) guide children's early language productions.

2.1.2. Animacy

The research also looks at noun animacy across all three constructions. For example, it includes the animacy of the theme and the recipient in the ditransitive construction, of cause in the periphrastic causative construction, and of the relativized noun in the relative construction. A simple division into inanimate and animate objects did not suffice for the analysis of animacy; in order to capture the finer details of the nouns in the constructions, a more sensitive gradation of animate objects was applied instead. Beside inanimate ones, nouns have been categorized as either typical animate nouns (e.g. *people, mom, dad*), names of people (e.g. *Adam, Bob*), animate toys (e.g. *lion, horsie*) and animate parts of animate beings (e.g. *hands, teeth*).[16] In some of the analyses, the categories of animacy were lumped together into two broader ones: inherent animates (names and typical animates) and contextual animates (animate toys and

[16] As far as pronouns are concerned, the pronoun *it* was classified as an inanimate, whereas other personal pronouns were classified as animates. For lack of a better solution, demonstrative pronouns (*this*, *that*, *these*, *those*) were classified as inanimates.

animate body parts). Based on previous works on the animacy hierarchy, several factors were taken into account in the creation of said categories. Aside from the taken-for-granted distinction between animates and inanimates, the sociocentric argument, empathy and definiteness play a role in distinguishing different types of animates (see Whaley, 1997, pp. 172-173). For example, an empathy-based argument has been crucial in ascribing toys (traditionally perceived as inanimate objects)[17] to the category of contextual animates, as (aside from representing living creatures) they invoke a sense of empathy in children as they represent more familiar items than some others might.

The analysis of animacy discussed in the construction-specific sections is type-based, which means it does not consider the token frequencies of particular items in their clause positions but rather focuses on data diversity. For example, the noun *mother* may have been mentioned 100 times more often than the noun *toy*, yet it is still counted as a single animate item. Moreover, it should be noted that the type-based animacy was calculated based on the number of different lemmas, not lexemes, thus increasing with the number of different inflected forms of the same lexeme rather than being lumped together under the same instance (e.g. *mum*, *mom*, *mommy* are counted as different types). This calculation, therefore, excludes the repetition of a single item and evaluates animacy based solely on the number of different items being relativized. A more detailed examination of token-based animacy, which takes into account the number of times each item has been used, is available in the Appendix 10.2.

2.1.3. Lexical richness and similarity

The measure of lexical richness (and enrichment for that matter) is certainly a good stepping stone in studying language acquisition. From the acquisition of first lexemes to more complex phrases, the obvious assumption (at least in Cognitive Linguistics) is that the syntactic development postdates the acquisition of basic lexicon. In any case, the description of early vocabulary, its progression through time and comparisons with adult language is a good beginning for any serious investigation of language acquisition.

[17] Note that only toys which exemplify what would otherwise constitute an animate being (if it were not a toy), such as animals, were labelled as animate toys.

Similarly to structural similarity, one of the ways in which we could determine whether children's language is truly and completely conservative[18] is to test the overlap in the use of particular lexical verbs in the studied constructions between age groups. If there is little or no novel content words used in the younger age groups in a particular construction, and if this choice of particular lexemes is heavily influenced by the frequency of the same in the received input (aka. adult language, and especially child-directed speech), it is arguably possible to deduce conclusions about children's early language learning mechanisms. In other words, much of the work presented will outline the most frequent tokens in various syntactic positions. So far, these approaches primarily looked at similarities in the context of verb-use, but this study will also include data on the particular nominals and their animacy.

Aside from lexical resemblance, the lexical exploration will also include various measures of lexical richness (or complexity). Depending on the approach and terminology, lexical richness can refer to lexical diversity (as in Daller et al., 2003), or it can encapsulate both lexical variation, number of errors, density, and related measures of sophistication (as in Read, 2000, p. 200-201). This thesis will use the term in the latter sense and will address some of the indices pertaining to density, variation and sophistication. In other words, lexical richness will be used as a cover term for aspects of both lexical density and diversity. While different measures and indices of lexical richness have been used a number of times in both first and second language acquisition and development (see Malvern et al., 2004 for an overview; also consult Richards et al., 2009), they have not always proven to be consistent in the evaluation of language development, and one of the aims of the research is to provide additional data on the matter.

One of the measures related to both lexical and structural development is the so-called "mean length of utterance" (MLU) which arguably gives an approximation of language proficiency, especially in child language. The measure can be obtained by dividing the number of morphemes by the number of utterances containing those morphemes, the usual number of which calculated for this purpose equals to 100. Other variations include dividing the number of syllables by the number of utterances, or the number of characters by the number of utterances. While it is not necessarily a direct indicator of lexical development, it can be

[18] In the context of language acquisition, conservative learning implies that a child does not go beyond what they have encountered in their immediate linguistic environment, firmly holding on to the use of constructions with specific lexical items as encountered in the input. In that sense, children's early language production is far from innovative with little to no differentiation between input and output.

interpreted as an indicator of lexico-syntactic development since both lexicon and syntax play a role in sheer length of utterance produced by children. The measure has often been used in early language development research, with various conclusions on its usefulness in ascertaining grammatical development solely on the basis of it. In a study on 87 children, ranging in age from 1;6 to 4;9, Blake et al. (1993) found that MLU qualifies for a valid measure for clausal complexity in the early language development stages, when the production encompasses single-word utterances, two-constituent clauses and clausal connectivity. However, MLU's sensitivity diminishes for increases in clausal complexity beyond this (complexity-driven) threshold. Devescovi et al. (2005) contributed to the discourse by emphasizing that both age and vocabulary significantly contribute to MLU, with vocabulary exerting a larger influence, particularly for cross-linguistic comparisons. Scarborough et al. (1986) proposed that MLU increases by approximately 1.2 morphemes per year up to the age of 5, with a declining growth rate post-42 months, while Jalilevand and Ebrahimipour (2014) further underscored the role of MLU as a general developmental indicator but cautioned against relying on it alone for determining the structural complexity of grammatical competence. For the purposes of this research, the program CLAN was used to calculate the MLUs for sentences containing the targeted constructions.[19]

However, in order to tackle the issue of lexical richness across different age groups with respect to constructions targeted by this research, several measures indicating their lexical rareness (sophistication), density and variation were explored. For this task, the research employed the LCA software (see Ai and Lu, 2010 & Lu, 2012) which covers 25 different measures of lexical richness and complexity.[20] There has been a great deal of research focusing on the best measures of lexical complexity and variety, and to this day the researchers tend to use different formulas depending on sample type, appropriateness of the measure, goal of the

[19] First, the sentences containing only the observed constructions were separated with the help of Excel tools and functions (it can be more difficult to isolate the final text after the last delimiter, but once equalized there is a combination of functions to solely isolate the final textual sequence in the row- e.g. if all the delimiters are recoded as question marks, the function TRIM(RIGHT(SUBSTITUTE(A1;"?";REPT(" ";100));100))). Second, using the CHAT conventions (see MacWhinney, 2018) the text is adapted for the use in CLAN program. Third, MOR grammar is installed and applied to the data (see MacWhinney, 2023b), so that morphosyntactic calculations can be conducted on it. Finally, the EVAL function was used, providing a number of valuable morphosyntactic measures of the input text. By the program's default, the words transcribed as 'xxx' (unintelligible or nonce words) were excluded from the MLU count.

[20] The LCA software is described as a tool that aims to automate the analysis of lexical complexity through 25 different measures of lexical density, sophistication and variation, relying on a system that has been fully elaborated in Lu (2012). The online version of LCA software *Web-based Lexical Complexity Analyzer* is available for the analysis of English language, incorporating both British and US varieties.

study etc. (see Malvern et al., 2004, Jarvis, 2013). Given that the matter of validity and interrelation of different measures concerning lexical complexity and variety is beyond the scope of this discussion, the analysis includes the data of measures which have thus far been proven as most reliable in lexical estimates of oral narratives (see Lu, 2012).

Lexical density (LD) has traditionally been used for the evaluation of different text types, whether transcripts of spoken language or written texts. The research has consistently shown that written texts tend to be considerably more content saturated than spoken texts (see Ure 1971; Halliday, 1985; Johansson, 2008), but apart from differences between spoken and written language, scholars have also looked at the linguistic and social contexts in which the texts have been produced or communication carried out. For example, the values have also proved to vary across different conversational situations, such as between job interviews and conversations between friends (see Zora & Johns-Lewis, 1989) or between direct face-to-face and tape-mediated English proficiency tests (see O'Loughlin, 1995). The index has been calculated by dividing the total number of content (lexical) words by the total number of words in the evaluated text.[21] The measure is used to evaluate how concentrated a particular text is in terms of ideas. It has also been called "an index of the degree of 'orality' versus 'literacy' in both spoken and written discourse" (O'Loughlin, 1995, p. 220). In the most straightforward sense, it represents an index of the relationship between lexical and grammatical items in the evaluated text. More 'literate' texts or speeches, such as those designed for formal occasions, are expected to have higher degrees of lexical density than informal ones. In the context of language acquisition and development, lexical density has been taken as particularly beneficial with studies in children with autism spectrum disorders (see Yoder and Stone, 2006; Yoder, 2006), where it was perceived as a reflection of productive vocabulary size and talkativeness (building upon Yoder et al., 1998). The index is valuable for understanding ratios between lexical and grammatical items in the text, where the latter are defined as function words limited in number (i.e. there is a finite number of them in a particular language). This approach was elaborated by Halliday (1985), who introduced an innovative modification to lexical density by taking into account the frequency of lexical words; if the lexical items were occurring frequently in the observed language in general, they would receive half of the value in the calculation

[21] Content words usually represent open-class as opposed to closed-class words, and in LD calculations, adverbs sometimes are the 'apple of discord'. Aside from including nouns, adjectives, verbs (beside modal and auxiliary ones) as lexical words in LD calculations, following Lu (2012, p. 192), the analysis in this thesis also includes adverbs with an adjectival base such as *fast* (can be used both as adverb and adjective) or *angrily* (adjectival base with the suffix *-ly*).

process, as opposed to those words which occur less frequently. Although this modification may provide more fine-grained results, the problem may arise with 'drawing the line' between highly frequent and lowly frequent lexical items. Following Engber (1995) and Lu (2012), the LD measure in this research is calculated as the ratio between content open-class words and the total number of words in a sample text.

Lexical sophistication is a measure that ties closely to lexical density because, much like density, it relies on the ratios between specific content words and, depending on different variations, the overall numbers of content words, the number of word types etc. In the analysis, in the thesis I will use the index of lexical sophistication as conceptualized by Linnarud (1986) or Hyltenstam (1988) that divides the number of sophisticated lexical words by the total number of lexical words. The index directly concerns the frequency of one's usage of lexical items that are not part of their narrowest and commonest language environment, but without delving into diversity. Although it may initially seem so, the basic measure itself does not reveal much about the range of lexical items used by the speaker. Theoretically speaking, it is possible for a person to have a poorly diverse vocabulary despite employing lexical items found beyond the most common ones, i.e. one could theoretically have good knowledge of the sophisticated lexicon without necessarily knowing the elementary terms. In this context, sophistication is to be regarded in theory as a separate concept from diversity, despite often corresponding to diversity in practice (for example, see Durrant & Brenchley, 2019). In this research, following Lu (2012), sophisticated lexical words were considered to be those not found amongst the most frequent 2000 words derived from the British National Corpus (Leech et al., 2014). The measures of sophistication have since been extended to specific parts of speech and have centred around verbs in particular. For example, one such measure has been suggested by Wolfe-Quintero et al. (1998), who proposed the measure of sophisticated verbs (CVS1) in the text to be altered along the lines of corrected type–token ratio (Carroll, 1964) so that the sample size effect is reduced. As opposed to just looking at the number of sophisticated verbs in a text, it takes into account word types; the index is calculated by dividing the number of sophisticated verb types by the square root of the doubled number of verbs in the sample text. Like with lexical sophistication, the goal is to assess how far the speaker goes beyond the most frequent words in the vocabulary, but with the focus on verbs, albeit with the introduction of diversity aspect.

As opposed to lexical density which concerns variation only in relation to parts of speech, lexical diversity provides an estimate of the range of lexical items used in the evaluated

text. High levels of lexical diversity would indicate more developed vocabulary levels, or at least a demonstration of the same. In order to achieve relatively high scores of lexical diversity, the evaluated corpora would have to represent speech or writing that employs different lexical items and does not repeat them that often in the text once they have been used. It is no surprise that, as lexical density, lexical diversity also tends to be of higher values in written texts as opposed to spoken ones. In the context of this work, the term is used as a cover term for the straightforward measures of range such as the number of different words in the text, as well as type-token ratios which look into variation with respect to size of the sample.

As Duran et al. (2004) note, the number of different words (NDW) is the most straightforward way to tackle lexical diversity in one's speech, but in order for it to work, the researcher has to standardize the size of the samples (for example, balance them either by the number of utterances or the number of tokens). For obvious reasons, the conclusions deduced from samples of different sizes are inherently wrong and defeat the purpose of the measure; if the size of one sample is bigger than the other, chances are that it will contain more different words. Both methods of sample standardization (either by the number of tokens or utterances) have their own advantages depending on which developmental aspects are emphasized, but the goal should be for researchers to agree on a single method so that cross-research comparisons could be carried out. Other methods that help with reducing sample size issues look at random subsets selected from the sample and take the mean number of types of words found across those subsets. One such measure used in this thesis (adopted from Lu, 2012) calculates the mean number of word types of 10-random 50 word samples (NDW-ER50). In both cases of NDW measures, the important thing to note is that they are to be understood as measures of ‘range’ (Malvern et al., 2004).

Type-token ratio (TTR) is in some sense a natural extension of NDW. The basic value is calculated by dividing the number of different words in a text by the number of tokens in a text. The term ‘type’ refers to unique words in a text, whereas ‘tokens’ represent each individual occurrence with a particular corpus position (i.e. each string of characters divided by space). In a text that contains three words, out of which two are repeated, the number of types will be two and the number of tokens three. Consequently, the measure is always expressed through values between 0 and 1, where the higher values represent greater diversity (Malvern et al., 2004). The obvious difference between TTR and NDW is that the former takes the size of the language sample into account, but the size can also become an issue. The problem is that the increase in

the number of tokens after a certain point becomes counterproductive for TTR values because maintaining lexical diversity in spite of the output volume produced would require extremely rich (or unlimited) vocabulary that is paired with little or no repetition. The shorter samples are more likely to (unjustly) benefit in comparison to longer ones. This is especially evident in the context of language acquisition and development, where it has already been stated and observed that linguistically more advanced children produce longer utterances (both verbally and morphologically richer) and they also produce them more frequently in given time frames (see Richards, 1987). The implication of the varying lexis being held constant is that TTR will negatively correlate with the number of tokens in the compared samples, and any increase in size will necessarily lead to a reduction in the value of TTR, only seemingly exemplifying a decrease in lexical diversity (ibid., p. 206).

One of the solutions for TTR being largely affected by the sample size is to calculate the average TTR based on the subset of samples derived from the observed sample. This so-called Mean segmental TTR method was proposed by Johnson (1944), who suggested that each of the samples gets divided into segments of the designated length and TTR calculated for each, with the final index being the result of average subsample TTRs. As Malvern et al. (2004) note, the measure is an equivalent of TTR for the expected number of words in the sample. Other transformations which aimed to achieve a constant value of TTR for samples regardless of their size include Root TTR (RTTR) by Guiraud (1960) and Corrected TTR (CTTR) by Carroll (1964). While the former divides the number of types by the square root of the number of tokens (RTTR), the latter does the same but by double the number of tokens (CTTR). The latest and perhaps the most precise solution has been the so-called Moving-average TTR (MATTR) by Covington & McFall (2010), also based on the mean TTRs, but with different method of capturing the subset, i.e. first a designated window length of words is chosen and then TTRs are computed for each of the ranges between the first and every subsequent word and the last (e.g. if the window length is 500, than MATTR is calculated by computing TTRs for 1-500, 2-500, 3-500 until the end), and the final value amounts to the mean TTR of the calculated subset TTRs. Since the compared samples in the research are already balanced for the random number of utterances (albeit not for the number of tokens), elemental values of type-token ratios such as MATTR and RTTR will suffice for the evaluation.[22] The exploration of this corpus data in terms of diversity is further supplemented by the transformed measure of verb variety SVV1

[22] For additional values on lexical richness, see the Appendix 10.2.

(as suggested by Wolfe-Quintero et al. (1998) following the original measure by Harley and King (1989)), which reduces sample-size effect. The idea behind the indices of verb-variation is the same as with lexical diversity, only instead of looking at the ratios between word types and their overall number, the indices take into account solely verb types and the total number of verbs in the sample.

Table 1. Measures of lexical richness considered in this study[23]

Index	**Formula**
Lexical Density **(LD)**	Nlex/N
Lexical Sophistication **(LS1)**	Nslex/Nlex
Corrected Verb Sophistication **(CVS1)**	Tsverb /$\sqrt{2\text{Nverb}}$
Number of Different Words **(NDW)**	T
NDW (expected random 50) **(NDW-ER50)**	Mean T of 10 random 50-word samples
Mean Segmental TTR **(MSTTR-50)**	Mean TTR of all 50-word segments
Root TTR **(RTTR)**	T/ $\sqrt{N}$
Corrected TTR **(CTTR)**	T/$\sqrt{2}$N
Squared Verb Variation **(SVV1)**	Tverb^2/Nverb

Legend: the number of tokens of words (N), lexical words (Nlex), sophisticated lexical words (Nslex), and verbs (Nverb), the number of types of words (T), sophisticated verbs (Tsverb), adjectives (Tadj)

In order to analyse lexical richness of the samples across age groups, in the thesis I cross-examined the average scores calculated on sentences containing the prototypical construction patterns explored in this research. The average scores for the listed measures were calculated on random samples of 200 utterances (exceeding 2000 tokens on average) for each observed age group.[24]

The analysis of lexical richness will be provided for utterances that centre around the observed constructions, and which consequently often extend beyond them. One reason for observing entire utterances and not isolated constructions (although there will be references to the same lexical indices pertaining solely to constructions as if they were isolated from the rest

[23] For additional explanation of indices, consult Lu (2012).

[24] The scores were rechecked with different database samples and they showed relative consistency in numbers. The choice of that sample size (random samples of 200 utterances) is additionally supported by the results of Zenker and Kyle (2021), who found many lexical diversity indices to be stable even on samples with less than 200 tokens, except for some measures of type-token ratios.

of the sentence) is because a more fine-grained description will already be provided in terms of head lexical elements and syntactic patterns that characterize these constructions. More importantly, it gives an information about the context in which the constructions are usually employed. A great part of analysis in language development and acquisition is sometimes too concentrated on the patterns as if they were isolated from the rest of the sentence just because there are no inherent constraints on their immediate linguistic environment, and yet many stable phrases (if not entire utterances) often go beyond the basic constructions. Furthermore, it may be interesting to observe how do entire utterances that contain these constructions differ amongst themselves, what role do larger language units such as sentences bear on the children's processing of the embedded constructions, and whether corpus-derived data would reveal something about the interplay of these structures in relation to age-differences.

Such approach to lexical richness in general is not something usually applied in language research; it is not common to observe measures of lexical richness for one specific construction (or the utterance that contains that specific construction, as is the case in this analysis) because the usual aim is to capture differences in view of density and diversity across samples that represent linguistic skills/competence in its entirety (whether written or spoken). Nevertheless, the lexical richness-based comparison of utterances that centres around one particular construction across age groups should additionally shed a light on some more specific construction-related developmental aspects. For example, the difference in lexical density demonstrated within a particular construction across age groups tells us whether the information uttered by these age groups in that particular construction is more concentrated. The increase in lexical diversity within a particular construction suggests that a particular age group is more lexically versatile in producing that construction. Moreover, given that three constructions are observed in this research, we might speculate as to why there might be a drop (or an increase) in the measure of lexical richness in one construction, as opposed to others. Finally, the measures can serve to confirm that the speculated differences in complexity between the observed constructions indeed exist when it comes to real language use, or even that a particular construction is more challenging to master from the perspective of language acquisition. The data may reveal that constructions themselves bear striking resemblances across age groups, but the question would remain whether the same goes for the rest of the language, even the segment which is closely related to the observed phenomena.

2.2. Research questions and hypotheses

So far, the introduction has outlined some of the key theoretical considerations for a corpus-based study that aims to tackle the issues surround language acquisition and development. In usage-based linguistics, the degree to which the input plays a role in language acquisition has been discussed and explored in a number of studies (see Section 1.2.3). The productivity of specific language patterns that uniquely intertwine with specific semantic features has been in the focus of Construction Grammar (see Section 1.2.2), while Cognitive Linguistics served to provide a more comprehensive theory of mind that allows for assertions about linguistic structures emerging from use (see Section 1.2.1). The research questions and hypotheses in the following subsections tie together the aforementioned theoretical and practical research in the hope of providing additional data on some issues relatively underinvestigated in this manner thus far.

2.2.1. Research questions

As stated in the introduction, the main goal of the thesis is to explore the structural and lexical similarities across the observed age groups (0-3, 4-6 and adults) using corpus-based analysis to target three specific constructions. Aside from examining inter-age and intra-constructional data, the study also addresses inter-constructional differences in the data retrieved, as well as the degree of 'item-based' character in each. The utilization of CHILDES in this manner is not entirely novel, with various studies, such as those by Liu et al. (2008), Goodman et al. (2008), Li & Fang (2011), and Shin (2022), having already employed CHILDES corpora to investigate parental input and child language, and having found significant relations between parental input and language produced by children in several respects (see Section 1.1). These studies predominantly concentrated on lexical aspects and similarities, whereas the contribution of this thesis is that it scrutinizes correspondence between specific construction slots and aspects associated to both lexical and structural complexity of the said constructions.

The research question pertaining to lexical and structural similarity builds upon usage-based accounts of language acquisition (see Section 1.2.3). Though some lexical and structural similarity is obviously expected between adult input and child output in any language

acquisition theory, the degree to which it occurs with specific sub-patterns of the constructions observed and the corresponding construction slots is a valuable contribution to the discussion on input-output correspondence. Furthermore, the research question pertaining to animacy builds on the idea that the imitative tendencies of early child language should affect the lexical character of items occupying the corresponding construction positions (cf. Buckle et al., 2017), but with a more dominant representation of specific animacy values which would otherwise be considered prototypical in the observed construction slot. The idea that children consider the semantics of items heard most frequently in the observed construction slots and consequently overproduce words which fit the dominant semantic frame is another extension of item-based suppositions about early language development that transposes to a semantic aspect such as animacy. Furthermore, the analysis of lexical richness and complexity has been used in various fields of linguistic research, with language acquisition being one of them (see Blake et al., 1993; Malvern et al., 2004; Yoder and Stone, 2006; Yoder, 2006; Li & Fang, 2011 etc.). However, the measures are usually employed on developmental data obtained longitudinally, whereas the inter-group comparison characteristic of this research has been underrepresented so far. Finally, the unique addition of this thesis concerns the inter-constructional comparison in terms of item-based tendencies, exploiting the idea that simpler constructions (see Section 1.4) should resemble the input more than others (cf. Tomasello & Brooks, 1999). The research questions and hypotheses that exemplify the aforementioned topics are the following:

1. Is there any lexical and structural similarity between age groups in terms of preferred lexical and syntactic choices in the observed constructions?

1a. To what extent are the constructions similar with respect to verb frequency?

1b. To what extent are the constructions similar with respect to noun frequency?

1c. Is there a difference between the observed age groups of children in terms of the degree to which they are similar? Does the 0-3 age group exhibit greater similarity to child-directed speech than the 4-6 age group?

2. Are there any differences in terms of animacy with respect to key lexical items (nouns) in the constructions?

3. Will the measures of lexical richness and complexity reflect the differences in language use across age groups with respect to the observed constructions?

4. Is there a difference across constructions in the degree of their item-based character?

2.2.2. Hypotheses

1. There will be a considerable lexical similarity in constructions across age groups in the observed constructions.

This assumption is anchored in the fact that children of the earliest age ought to reproduce the most frequent patterns as they occur in their immediate linguistic environment. Having in mind the item-based tendencies in early language, the following hypotheses are then derived from this:

1.1. The selection of certain verbs in constructions will reveal a considerable similarity across age groups.

1.2. The selection of certain nouns in constructions will reveal a considerable similarity across age groups (albeit smaller than when it comes to verbs).

1.3. In comparison to other two, the youngest age group's production of constructions will be characterized by a greater disproportion between the most frequent items/patterns and the rest (most frequent ones will be more pronounced than in other age groups).

1.4. The language of 0-3 age groups will be more similar to child-directed adult speech than the language of age groups 4-6 when it comes to the ratio of the use of certain structures and lexemes.

As children grow up, their language should be more similar to the adult one in terms of how competent and productive they are, but not in terms of reiterating lexical choices that appear in specific construction slots or even the degree to which they reproduce the same patterns in the input. If this assumption turns out to be correct, in a way it agrees with the following: child-directed speech does help at the earliest ages, but its reach is limited in the sense that children learn language imitatively up to a certain age, but as their cognitive systems develop, they begin to rely on other mechanisms such as distributional/statistical learning and 'pattern finding' skills. At that moment, the speech aimed at children becomes less relevant, and other spontaneous speech takes precedence, from which the slightly older children begin to derive their own conclusions about what language use allows and disallows in terms of context, grammaticality, logic etc. This fact may additionally be supported by more equivalent ratios in children and

adults regarding certain lexemes and structures, as well as a simultaneous explosion in terms of the frequency of constructions after reaching the age of four.

2. The language of 0-3 age groups should be least flexible when it comes to the animacy of key lexical nouns in the constructions; meaning that the category of animacy considered prototypical for particular lexemes in the observed constructions will also be the most saturated in the youngest age groups when compared to others.

This hypothesis reflects the idea that the children's learning of constructions is not just imitative in character, but also distributional; that is, by taking into account the linguistic context and remembering various instances in which a particular item (or a phrase) can be used, children should be able to infer on some lexical properties of these items, such as animacy. In turn, their early suppositions concerning the use of these items (along with imitative tendencies) should reflect on the character of these items; they will favour the prototypical (most frequent) uses found in the input. In other words, there is a reason to assume that children's representation of specific language patterns closely ties to the semantic character of the same construction which they encounter most often in the input (for experimental data, see Buckle et al., 2017).

3. The measures of lexical richness and complexity will differ across age groups despite the potential lexical similarity.

Despite the item-based character of language development, the measures of lexical richness and similarity should still reflect differences in the syntactic and lexical elaborateness of older age groups in comparison to younger ones.

4. Simpler constructions will be more item-based in character than more complex constructions; that is, ditransitive constructions should be more item-based than causative constructions, and causative constructions should be more item-based than relative constructions.

This prediction builds on the perspective explained in Tomasello & Brooks (1999), which assumes that the learning of the earliest constructions does not take place in the context of abstract syntactic patterns that children have access to from an early age, but rests on the emulative learning of constructions around its central words. In this sense,

it could be expected that the learning of patterns that justify the label of 'construction' to a greater degree (because they have a more stable syntax of an almost inherent meaning; e.g. ditransitives compared to relatives) will be anchored to their central elements to a higher degree as well. On the other hand, constructions with greater 'combinatorial' potential in children should be less similar to the input itself.

2.3. Method

2.3.1. Approaching the corpus

In this construction-aimed corpus-based research, a key aspect of the methodology involves operationalizing the constructions we are examining. In this sense, each section that looks into the data on particular constructions also includes an introduction covering the theoretical background of the phenomenon that is going to be observed, the overview of previous research, and the operationalization of the observed phenomena. Essentially, the undertaken research process can be outlined in the following way: the so called "scouting" phase which included characterizing the phenomenon and reviewing the relevant literature, followed by the observation of the linguistic phenomenon in the primary sources, and finally the gathering of new information and applying deductive reasoning (as described by Gries, 2013, pp. 8-9). In other words, the more general theoretical overview provided in Chapter 1 is complemented with construction-specific research overview that also serves as a background for the proper methodology (the extraction of the constructions from the corpus) – the sections operationalizing individual linguistic phenomena observed serve as introductions to chapters analysing them (see Section 3.1, Section 4.1, and Section 5.1).

The most time-consuming process in this research was the data curation process, i.e. extraction of data from the corpora and the removal of "false positives" from the final dataset. In order to retrieve the target constructions from the corpus, it was first necessary to locate them in the transcriptions (enabled by the Sketch Engine-integrated part-of-speech automated tracking tools) and to filter out those instances that actually represent the constructions targeted by the research from those that, for one reason or another, manifest the same part-of-speech pattern but do not qualify for said constructions in terms of meaning or even structure (for example, due to repetitions, interjections, redundancies or non-registered sentence breaks, an isolated syntactic layout of the speech segment may accidentally correspond to target

construction, thus making it a false positive). Finally, to arrive at the total number of 78 871 instantiations of the constructions produced by the speakers of the examined age groups in the corpora (i.e., the raw frequency of occurrence for all three constructions combined), a total of 345 753 occurrences had to be hand-filtered following the automated phase of dataset extraction. In other words, the false positives comprised a total of 77% of all instances initially located in the corpora according to the syntax outlined in Appendix 10.1.

Corpus linguistics has often been dubbed "a distributional discipline", owing its name to the logic of offering answers to questions with regards to the frequency of occurrence and location of something in the corpus, as well as the frequency of co-occurrence of linguistic expressions and the way in which the linguistic elements are used in different contexts (Gries & Newman, 2013, p. 274). Two variables fundamental to this research cover language development and age. Language development is a dependent variable and is measured indirectly through examination of structural characteristics and lexical enrichment. Age, which in this study, based on the age ranges available in CHILDES, takes one out of three values (0-3, 4-6 and 18+), is a common independent variable in linguistic research, not only in first but also in second language acquisition (Rasinger, 2013, pp. 22-23). For this reason, the analysis mostly delves into frequencies of particular patterns and cross examines them for the said age groups. Two types of data can be extracted from corpora such as CHILDES: word frequencies or construction frequencies and their variability, and second, frequencies and distributions of co-occurrences of different linguistic elements. While most of the discussion delves into the quantitative data on frequencies and distribution, the qualitative observations of the collected data reflect upon the establishment of word meanings in the context of language acquisition and development,[25] as well as the role of specific lexical choices in their respective constructions.

2.3.2. Sample

A part of the methodological work for this project has already been done in the sense that the study is based on an already existent family of corpora. In other words, the strength of any imaginable generalisations that might be deduced in this endeavour, depend partly on the methodological soundness of the data collection process, i.e. it is important to bear in mind for

[25] Corpus semantics literature suggests that the words acquire meaning precisely through co-occurrence with other words, especially when these co-occurrences are frequent (Corrigan, 2012, pp. 271-285; also see Hanks, 2000).

what reasons and how well the data has been collected. While data collection remained outside of my control, it is necessary to at least audit the methodological procedures used by the linguists who gathered the data that now belongs to the CHILDES corpus.

CHILDES is nowadays described as an "arranged database of transcriptions of spoken language contributed by researchers who originally collected the data to conduct their own research" (Corrigan, 2012).[26] Nowadays, the corpus contains a number of different transcriptions from monolingual speakers belonging to different language groups, whether Romance, Germanic, Slavic or English (MacWhinney, 2015). A relatively small part of the corpora is expanded by narrative data, bilingual and linguistically impaired speakers, and these were eliminated before the elicitation of linguistic phenomena continued. It is important to note that while CHILDES contains the age group 0-3, most of the data on early speech does not capture the stage of infancy (in its narrower etymological sense the term corresponds to the age at which children do not speak), which lasts up to one year of age. This absence is not necessarily a byproduct of inconsequentiality of this age span on the language development, but has to do more with the fact that children only start to provide some degree of sensible output after they have turned one year of age. Indeed, some linguists have argued that the age of infancy bears a quite important role on the acquisition of language (Fenell, 2012, pp. 3-17), but the focus of this research is on the preschool years, which manifest most development in spoken language, despite the arguments of some linguists that some structures, such as relative clauses and passives, appear in later stages of language acquisition (Bloomer & Wray, 2006, p. 32).

Nonetheless, the main advantage of CHILDES in this type of research is not only that it provides the transcriptions of spontaneous speech between children and parents, but also its sheer volume. The size of the English corpora, which exceeds 22 million tokens, suffices to make trustworthy generalisations on the data. Sometimes it is important to decide on the size of the sample used prior to the conducting of the research. This can especially prove problematic if one generates large amounts of data only to find out that the process was unnecessary, or worse yet, that the size of the data generated is not nearly enough to have statistical power for making desired generalisations (Rowe, 2012, p. 199). The need for such sample sizes is also

[26] Conversations between children and adults comprise most of the database and this is coded in a standardized format called CHAT (Codes for the Human Analysis of Transcripts). These transcripts are then stored as separate files within CHILDES, under the names of the researchers who did the work. While the knowledge of CHAT notation aids the investigation of the CHILDES corpora, the extraction process of desired linguistic phenomena has been simplified by the existence of corpus managers and software such as Sketch Engine.

one of the reasons why quantitatively ambitious studies in this research field require a considerable amount of funding, abundance of researchers and an enormous amount of time. Thanks to the number of previous language acquisition studies and the help of necessary computer tools, we are now able to use this data to reach conclusions that might help us gain better insight in the mechanisms behind language acquisition.

3. DITRANSITIVE CONSTRUCTION: CORPUS-BASED COMPARISON OF CHILD AND ADULT SPEECH

The following subsections of the chapter provide the theoretical background on ditransitive constructions, including the operationalization of the syntactic pattern and the previous research on the matter, especially in the context of language acquisition and development (Section 3.1). The data retrieved, the results of the analysis, and the interpretation, are introduced immediately afterwards (Section 3.2).

3.1. Operationalization and previous research

3.1.1. Defining ditransitives

Ditransitive constructions are often defined in relation to intransitive and (mono)transitive constructions. As the name suggests, ditransitive constructions double in transitivity or, in other words, they tend to take two objects rather than one – which is why they are also dubbed double-object constructions. For example, the verb *read* can take on the intransitive, monotransitive and ditransitive form (here, we also need to be careful and distinguish between ditransitive constructions and ditransitive verbs). For instance, in the sentence *She read for a while*, the verb *read* is an intransitive one, in the sentence *She read the newspaper* it is a monotransitive and in *She read us a story*, it is a ditransitive one (see Huddleston & Pullum, 2002, p. 53).

Aside from transitivity, the verbs also differ depending on the number of arguments the predicate controls, that is, they can be categorized according to valency. Although transitivity correlates with the verb's valency, there are differences depending on whether the complement in question is a core or a non-core one (a non-core complement is similar to adjunct, but is not to be confused with it – see Bergs, 2020). Examples from Huddleston & Pullum (2002, p. 219) demonstrate the differences between different predicate structures in English:

(1) (a) Intransitive and monovalent - *He died.*
(b) Intransitive and bivalent - *This depends on the price.*
(c) Complex intransitve and bivalent - *Ed became angry.*
(d) Monotransitive and bivalent - *He read the paper.*
(e) Monotransitive and trivalent - *He blamed me for the delay.*

(f) Complex monotransitive and trivalent - *This made Ed angry*.

(g) Ditransitive and trivalent - *She gave him some food*.

(h) Ditransitive and quadrivalent - *I'll trade you this bicycle for your binoculars*.

This study looks into ditransitive, trivalent, and quadrivalent (which tend to be rare) structures. The first commonality for the ditransitive construction is that it takes two objects, a direct object, which is more directly affected by the predicate (it receives or undergoes the action of the verb), and the indirect one, whose nuclear function is to name an end-point, goal or a person who benefits (see Fish, 1968, p. 863). The canonical ditransitive construction exhibits both objects overtly; however, in some cases the indirect object can be omitted when the context is clear to the parties involved (especially with verbs *show, send, tell, pay, lend, offer* etc.). For instance, the conversational English might allow sentences such as *He showed (us) his muscles* or *He sent (them) a cake for their birthday*. Typically, the omission of an indirect object occurs either when it is irrelevant to named the indirect object, or when it is possible to deduce the indirect object from the context (Carnerero, 2007). When the indirect object is omitted from the ditransitive construction, the derived pattern is called "dimonotransitive" (ibid.).

When it comes to ditransitives and their monotransitive prepositional variations, five different verb classes have been distinguished based on the constructions they license: (2a & 2b) can license solely the canonical ditransitive, (2c & 2d) can license the canonical ditransitive and the monotransitive with *to* phrase, (2e & 2f) can license the canonical ditransitive and the monotransitive with *for* phrase, (2g & 2i) can license solely the monotransitive with *to* phrase and (2j & 2k) can license solely the monotransitive with *for* phrase (examples from Huddleston & Pullum, 2002, p. 309):

(2)

(a) *I spared her the trouble*. (b) **I spared the trouble to/for her*. [indirect object only]

(c) *I gave her the key*. (d) *I gave the key to her*. [indirect object or *to*]

(e) *I bought her a hat*. (f) *I bought a hat for her*. [indirect object or *for*]

(g) **I explained her the problem*. (i) *I explained the problem to her*. [*to* only]

(j) **I borrowed her the money*. (k) *I borrowed the money for her*. [*for* only]

*ungrammatical

According to Huddleston & Pullum (ibid.), the first column represents canonical ditransitive verbs, while the second represents monotransitive verbs followed by direct objects and non-core complements. In terms of constructions, example (2c) represents the canonical ditransitive construction, while (2d) represents its dative alternation.

The aspect of 'transfer' has been viewed as a common denominator for the majority of ditransitive constructions. Typically, the agent initiates the transfer of a particular theme to a particular recipient, and this transfer most often has a possessive component; for instance, with the verbs *give, lend, award, deliver* etc. (Haspelmath, 2015, p. 20). This aspect of 'transfer' is not always literal and it can vary across constructions in terms of abstractness. For instance, verbs such as *offer* and *promise* do not necessarily entail a transfer of a physical object to a particular person (e.g. *offer someone your insight*), although such examples frequently occur with them (e.g. *offer someone a sandwich*). Likewise, despite the fact that the possessive aspect is implied, it does not necessarily have to be realized. Verbs such as *show, teach* and *tell* are even further on the spectrum of abstraction, but they are also normally included in the ditransitive domain; the term 'cognitive transfer verbs' (Haspelmath, 2015, p. 20) has been used to denote such examples where a transfer typically results in some sort of enrichment (whether positive or negative) – the recipient is enriched by whatever they have been shown, taught or told.

Both the theme and the recipient in ditransitive construction tend to demonstrate different behaviour patterns in terms of syntax and semantics. Thus, in English we find ditransitive alternations (a double-object ditransitive and a prepositional dative) which are argued to display differences not just on the syntactic level, but on the semantic one as well. Syntactically speaking, the difference is obvious in the structure of the predicate, which primarily diverges in the order of object arguments between the two. Despite the superficial semantic alikeness, it has been argued that the prepositional alternation changes the meaning as well (see Gropen et al., 1989, Goldberg, 1992, Harley, 2002 etc.). For instance, it is possible to argue that in the example *Vladimir sent his comrades the message*, the comrades eventually came into the possession of the message in question, whereas in the example *Vladimir sent the message to his comrades*, although the comrades were sent the message, it is not entailed that they received/accepted it. In other words, the double-object construction in English language is said to express a caused possession (A causes R to have T), while the prepositional dative construction conveys a caused motion (A causes T to move to R) (Haspelmath, 2015, p. 26).

Analogous semantic alternations have been observed on ditransitive construction denoting less overt ‘transfers’, such as with cognitive transfer verbs:

(3) (a) *John taught the students French*
(b) *John taught French to the students*

Example (a) entails that the students have mastered French, while in the example (3b), it is only obvious that the students were indeed educated, but it is not conclusive that his process was successful in the end (Green 1974). Similarly, ditransitive alternations with the verb *throw* also vary in interpretation:

(4) (a) *Dick threw John a ball*
(b) *Dick threw a ball to John*

In sentence (4a), it is implied that John has caught the ball he was thrown. However, in the second example, John is perceived more as a target than a recipient; i.e. in the prepositional dative alternation, John could have easily been dead or asleep while being thrown a ball, whereas this is not conceivable in the first sentence (Pinker, 2013, pp. 97-98). Furthermore, a distinction between caused motion and caused possession was explored on the verbs *send* and *throw* by Rappaport Hovav and Levin (2008), who argued that both may take either of the aforementioned meanings (either the recipient has come into the possession of the theme or they have not). Such distinction is important, because it not only allows us to distinguish between the two types of construction, but it also make it easier to define and operationalize a ‘prototypical’ ditransitive, which we can then define as only the one that takes an implied ‘possessor’ as its recipient.

In many cases, the meaning of a direct object is closely related to that of a recipient. This was noted by Larson (1988) on the prepositional dative alternation but the same logic applies to double-object constructions. Consider the following:

(5) (a) *Beethoven gave the world the Fifth Symphony,*
(b) *Beethoven gave his patron the Fifth Symphony*

In (5a) we assume that the “transfer of possession is metaphorical”, and the *Fifth Symphony* refers to a composition in general. In sentence (5b), the implication is that *Beethoven* gave a physical copy where the composition is written to his *patron*. In other words, the nature of the

recipient governs the exact semantic role that can be assigned to the direct object in a ditransitive construction.

Goldberg (2002) has argued that that the ditransitive construction needs to be distinguished from other prepositional alternations that share similar semantics, i.e. she considered it wrong to view ditransitives as derivations from its "corresponding paraphrases", and instead suggested that we treat these as constructions in and for themselves (ibid., p. 336). In other words, she calls the 'double-object construction' a ditransitive construction, while the dative construction with the preposition *to* is to be called a 'caused-motion construction' (e.g. *Jack gave a gift to John*), and the prepositional 'for alternation' should be named something along the lines of 'transitive construction + benefactive adjunct construction' (e.g. *Jack bought a gift for John*).

When it comes to the complexity of arguments, it has been noted that the canonical double object ditransitive favours 'simpler' recipients (or generally indirect objects) when compared to its dative alternation. For example, speaker will tend to choose a construction such as *He gave his son a couple of CDs* more frequently than *He gave a couple of CDs to his son*, whereas when it comes to more complex recipients, the preferred construction is the one with the prepositional complement – e.g. *He gave the spare copy to one of his colleagues* is likely to be preferred over *He gave one of his colleagues the spare copy*. Correspondingly, the construction with a pronoun as a direct object (e.g. *I gave Kim it*) seems awkward or even grammatically 'inadmissible' for English language speakers (examples from Huddleston & Pullum, 2002, pp. 309-310).

In terms of semantics, some scholars have argued that the ditransitive construction shares some features with the causative one. For instance, the verb *give* in a ditransitive construction can supposedly be decomposed into 'cause-to-have' predicate (see Richards, 2001; Harley, 2002; Jung & Miyagawa, 2004). In other words, many of the double-object patterns in English language can be decomposed into a causative predicate.

Another restriction on the double-object ditransitive construction has been observed in terms of argument animacy (see Gropen et al., 1989; Harley, 2002), which suggests that the referent in the first object position (indirect object in a double-object construction) must be an animate one. The following examples serve to illustrate this claim (taken from Harley, 2002, p. 35):

(6) (a) *The editor sent the article to Sue.*

(b) *The editor sent the article to Philadelphia.*

(c) *The editor sent Sue the article.*

(d) *The editor sent Philadelphia the article.*

While it is harder to observe how examples (6a) and (6c) diverge in interpretation, the nature of the noun *Philadelphia* calls for a different reading between a double-object construction and the prepositional dative variant. The prepositional dative allows for an interpretation where *Philadelphia* is not necessarily an animate being, but more of a location, whereas this type of interpretative flexibility is not allowed in the double-object construction. In a double-object construction, *Philadelphia* stands for some sort of organization (a group of people) and as such, represents an animate recipient. This rule derives from a semantic criterion stating that the referent in the first object position in a double-object ditransitive must be a "prospective possessor of the referent of the second object" (Gropen et al., 1989, p. 207), and considering that a prospective possessor is perceived as an animate being, the conclusion is that only animate recipients occur in double-object ditransitive constructions. On the other hand, the prepositional dative allows for the indirect object to be interpreted as a location and not necessarily a possessor or a recipient. Indeed, there is a lot of variability depending on the choice of verbs in a given construction. For instance, it has been observed that with verbs such as *show, give* and *tell*, the theme tends to be an inanimate object, while the recipient is consistently animate. On the other hand, verbs such as *bring*, *send* and *throw* allow for the recipient to be an inanimate thing, which is in line with the interpretation that these verbs also more readily accept more 'goal-like', as opposed to always requiring 'possessor-like' recipients (Haspelmath, 2015, p. 37).

3.1.2. Ditransitive verbs vs. ditransitive constructions

The constructional approaches to ditransitive constructions treat this syntactic pattern as a symbolic unit, meaning that there is a deeper semantic characterization almost inherent to this particular argument structure construction. This is an interpretation that moves away from lexically-rooted perspectives, where the verb in a construction dictates the meaning of it, i.e. every variation in the meaning between these sequences is explained primarily via verb senses.

This dispute relates to one's outlook on a range of meanings that can be attributed to a particular construction; the question is whether ditransitive constructions cover a wide range of similar but nevertheless distinct meanings, or whether these meanings are enough alike to be treated as one. In other words, are these semantic alterations explicated by verb senses, or do we attribute the construction one underlying meaning that encompasses all semantic realizations for each of the instantiations. For instance, the following examples illustrate different semantic entailments realized in the same construction (examples from Gisborne & Trousdale, 2008):

(7) (a) *Jane gave Peter a cake.*
(b) *Jane sent Peter a cake.*
(c) *Jane faxed Peter a letter.*

The first sentence (7a) entails that the Peter has received the cake, while this is not the case in second sentence (7b), where it is only understood that the cake was indeed sent, but that it did not necessarily reach its destination. In the third sentence (7c), Peter also has not received the physical copy of the letter, but was instead sent only the contents of it. In some cases, the differences in interpretation seem to exist quite clearly and it is necessary to recognize them; however, aside from recognizing these different semantic entailments, we may also recognize the commonalities. Instead of arguing that the verb 'send' has multiple word senses that depend on the syntactic frame which it occupies, it is possible to argue that there is an underlying meaning inherent to the construction, in which we can fit a more general "underspecified sense" of the verb *send* (Gisborne & Trousdale, 2008, pp. 2-3). Therefore, in the constructional approach, there is a meaning that can be associated with the construction itself to the point where even if you fit nonce words in it, there is a high probability that the speaker could infer the semantic constraints of the syntactic pattern in question.

In order to distinguish between ditransitive verbs and ditransitive constructions (at least in terms of how they are conceived in Construction Grammar), we have to recognize the fact that some transitive verbs and intransitive verbs can appear in a ditransitive pattern, as well as the fact that ditransitive verbs can occur in other types of syntactic environment. For instance, a ditransitive verb such as *send* can take on the intransitive form with a prepositional complement (e.g. *He sent for him*) or in a transitive prepositional pattern (e.g. *He sent him into the comma*). This is not exclusive to the verb 'send' but it also happens with prototypical ditransitives such as *give* (consider the intransitive prepositional construction such as *Mary*

gave freely to the poor). More importantly, verbs which are not usually perceived as ditransitive verbs occur in ditransitive constructions in the context of 'transfer' (e.g. intransitive *blow* in *Mary blew John a kiss* or a transitive *throw* in *Chris threw Pat the ball*) (examples from Stefanowitsch & Gries, 2003, p. 227). Examples such as these may be treated as evidence for the existence of inherent meaning to a ditransitive syntax.

In the example *He gave it all he could***,** there is no real recipient, at least not one that we could locate in a physical word, i.e. the 'recipient' in this case is typically something abstract, in a form of a goal that the agent has to attain. One could argue the construction to be idiomatic, but the idea of 'transfer' still exist to a certain degree – in order to attain a goal, a person will sacrifice their time and will, and thus in the process, the attainment of a particular goal becomes the recipient of a person's effort. Clearly, there are examples in English where a ditransitive construction does not convey the usual thematic roles of *agent-recipient-theme*, but from a comprehensive viewpoint, the interpretation of the said example falls under the scope of the semantic extension allowed by the ditransitive construction itself. Furthermore, in the aforementioned example, the theme is represented by a clause, but it may be replaced by a noun phrase at any time. In fact, the construction is rather productive; apart from the clausal complement, we can use an adjectival one (e.g. *He gave it his best*) or a noun phrase instead (e.g. *He gave it a shot*).

Trousdale (2008) discusses composite predicates that resemble ditransitive constructions in terms of lexicalization and grammaticalization, i.e. according to him, the construction as a whole can also be subject to lexicalization and grammaticalization processes. This interpretation is supported by the construction-based account and it can be observed on the examples of constructions resembling the canonical ditransitive construction. Most *give* composite predicates share the same semantic and syntactic properties; a 'give-gerund' composite predicate (e.g. *I gave him a kicking*) resembles the 'give ditransitive construction' in many aspects, but the partial idiomaticity of the expression can sometimes dissociate it from the canonical ditransitive. For instance, *I'll give her a seeing to* is even more idiomatic, in that it typically means *I'll have sex with her* or *I'll physically harm her*, and only the second meaning is evident from its prepositional variant (*I'll see to her*) (examples taken from Trousdale 2008; 35). The point is that not all 'give composite predicates' entail some type of transfer regardless of their shared syntactic characteristics, and part of the reasons for this are the processes of grammaticalization and lexicalization, and this does not necessarily conflict with the

construction-based account and the assumed shared semantic alikeness of particular grammatical patterns (the verb *give* developing a telic aspect in the 'give-gerund' composite predicate is interpreted here as an example of grammaticalization, while its idiomatic meaning is viewed as an example of a lexicalization process) (ibid., pp. 35-36) .

Likewise, some reframing effects of the verb *give*, and consequently some ditransitive forms, have been observed in English. The verb *give* in a ditransitive construction is said to influence the semantics of the expression it paraphrases. For instance, the expressions *I kissed Gwen* and *I gave Gwen a kiss* differ in that the verb *give* in the ditransitive construction turned an atelic event into a telic one; the first expression does not imply any temporal boundaries, while the second has the connotation of the event taking place instantaneously (Trousdale, 2008, p. 40).

Although this research adopts the Construction Grammar's perspective on constructions, this does not imply a rigorous 'structure-above all' approach to the studied constructions, or the children's language for that matter. As Goldberg stated herself, despite the fact that we observe shared properties of certain linguistic pattern, we still have to acknowledge the existence of same patterns with independent meanings, which means that we cannot go completely beyond the verbs as such in our interpretation of these structures:

> "That is, there do exist instances of constructional ambiguity: a single surface form having unrelated meanings. It must be emphasized that it is not being claimed that meaning is simply read off surface form. What is being suggested here is simply that by putting aside rough paraphrases and considering all instances with a formal and semantic similarity, broader generalizations can be attained. In order to identify which argument structure construction is involved in cases of constructional ambiguity, attention must be paid to individual verb classes. In fact, in order to arrive at a full interpretation of any clause, the meaning of the main verb and the individual arguments must be taken into account." (Goldberg, 2002, p. 335)

When it comes to verbs used in ditransitive construction, the collostructional analysis[27] done on the International Corpus of English (ICE-GB) revealed that the verb *give* is by far the most

[27] Collostructional analysis is a set of quantitative corpus-linguistic methods enabling researchers to quantify the strength of the relationship between word constructions and the grammatical structures they appear in. The primary goal is to identify the lexical items that commonly collocate with a specific grammatical construction, which in

frequent one (Stefanowitsch & Gries, 2003, p. 229). Other salient collocates were also *tell, send, offer, show, cost, teach*, followed by *award, allow, lend, deny, owe, promise, earn, grant, allocate, wish, accord, pay, hand, guarantee, buy, assign, charge, cause, ask, afford, cook, spare, drop*. In terms of Construction Grammar, such analysis might suggest a connection between the inherent meaning of construction and the usually assumed meaning of the collocate verb. In other words, the semantic properties of the most significant collexems tend to correlate with the perception of what constitutes a ditransitive verb, i.e. the more salient the verb used in the ditransitive construction, the more likely it is for speakers to conceive it as a ditransitive verb (ibid., p. 229). If anything, lexical analysis of frequent collocates allows us to claim that a ditransitive construction with the verb *give* can be regarded as a prototypical one, or, from the perspective of Construction Grammar, that the speakers most frequently associate it with the meaning and syntax of the ditransitive construction.

The dominance of the double-object ditransitive construction over the prepositional dative alternation was observed by a number of studies and authors (see Biber et al., 1999; Campbell & Tomasello, 2001; Stefanowitsch & Gries, 2003; Siewierska & Hollmann, 2007 etc.). A corpus-based research on the Lancashire dialect texts (that constitute a part of the British National Corpus) has revealed the vast majority of ditransitive clauses to be double object ones (a total of 83%) and featuring the following verbs: *blow, bring, buy, cook, draw, get, fetch, give, hire, make, offer, owe, pay, post, read, save, send, sell, show, take, teach, tell* and *write* (Siewierska & Hollmann, 2007).

3.1.3. Previous research

When it comes to language acquisition research, ditransitive constructions represent an interesting object of study due to a number of reasons, some relating to the complexification of simpler structures (such as transitives) and their developmental trajectory in child language, and the other relating to divergent interpretations and operationalizations of this syntactic pattern (e.g. the 'lexical rules' account vs. the constructivist approaches). Different suppositions regarding the interplay between syntax and semantics, how they are decoded and explicated,

turn facilitates semantic analysis of the grammatical constructions observed (see Stefanowitsch, 2013 and Hilpert, 2014).

present important concerns to language acquisition researchers for the same reason language processing research relates to the field of language acquisition (see Gropen et al., 1989; Snyder & Stromswold, 1997; Campbell & Tomasello, 2001; Goldberg, 2006). The research in first language acquisition has found ditransitive constructions interesting for a number of reasons, the most important ones being the fact that they are acquired and demonstrated relatively early in children's speech, that they share relatively stable semantic properties (some sort of 'transfer' being the recurrent theme) and that they are cognitively complex (the action they denote involves three participants) (Dixon, 1991).

Many studies that have dealt with the acquisition of the ditransitive construction have looked into the order in which datives are acquired. The results of these studies have consistently shown that the acquisition of (ditransitive) double-object construction arises earlier in children's speech than the prepositional 'to/for' dative alternations (see Bowerman, 1990; Pinker, 1984; Snyder and Stromswold, 1997; Campbell and Tomasello, 2001). The argument concerning the order of acquisition, which has also been the subject of research in second language acquisition, has been used to support the claims of what constitutes a derived construction and which formal pattern can be regarded as the 'source' (see Sánchez-Calderón & Fernández-Fuertes, 2018); for instance, if double-object constructions are performed earlier than 'to/for' datives, one could use this finding as a part of the argument for the claim that 'to/for' datives are in fact the ones that are derived (and not the other way around).

The research on the acquisition of ditransitive construction has revealed that children master such patterns at a relatively early age – the age range at which the children master the double-object construction tends to vary between 1;8 (years; months) to 2;11, whereas when it comes to the prepositional alternation or the 'caused-motion construction', children tend to master it sometime between 2;0 and 3;4 years of age. In some cases, they master these constructions simultaneously, but the tendency seems to be that the prepositional pattern is acquired later, and the temporal gap between the production of these constructions in early language may be as long as 12 months (Snyder & Stromswold, 1997). The hypothesis that children produce double-object constructions before the prepositional *to*-dative was corroborated by Campbell and Tomasello (2001), although the overall age at which the children acquire datives was later than posited before; for example, one child first produced a dative construction with 2:9 years of age. Their conclusion with regards to the age of acquisition was

that the strong correlation between the points at which children start to produce these two constructions suggests a grammatical relatedness between the two.

One of the first major papers on the acquisition of datives was supplied by Gropen et al. in 1989, where they argued against the claims that input plays a major role in children's acquisition of dative constructions. Similarly to this one, an important finding that raised a lot of questions regarding the acquisition of ditransitive construction was the one by Snyder and Stromswold (1997), whose comparison of input and output led them to infer that the frequency of certain ditransitive patterns in the input children receive did not play a role in terms of their output. Although their focus was on the age of acquisition and whether the increased frequency of ditransitive constructions with the verb *give* might facilitate the process of acquisition, the absence of correlation between the recurring patterns in input and the age of acquisition was indicative of the input's overall role in the early language development. Such findings were challenged by Campbell and Tomasello (2001, p. 257), who also expressed some methodological concerns regarding the previous works, arguing that it would be far more beneficial to approach the matter by numerically separating instantiations according to the verbs used and applying a cross-examination of dative-construction frequencies between child and adult language. Their results pointed in a different direction to that of previous research, i.e. the general similarity between the frequency of use when it comes to specific verbs revealed a strong correlation between the adult input and children's output. The data was interpreted within the framework of usage-based models, which predict exactly that the verbs heard most often by the children would reflect on their language use, and in turn produce similar tendencies in terms of frequency of use– children did, indeed, produce most frequently those verbs which they heard most frequently. Moreover, this lead them to believe that children indeed learn double-object dative first precisely because it was the construction they tended to hear most often in their immediate linguistic environment.

As stated, Gropen et al. (1989) argued that the frequency of a particular construction in the input (the language which they hear from adults in their immediate environment) does not affect children's acquisition of the double-object dative. Their analysis was also conducted on the data from CHILDES, or more specifically, on the data from Brown's and MacWhninney's corpora. Their results suggested that neither of ditransitive alternations (double-object and the prepositional dative) consistently emerged first in the acquisition, despite suppositions that one precedes the other in the acquisition order. In the end, they reported that "about 95% of the

children's double-object sentences (tokens) and about 86% of the verbs (types) they use in double-object sentences could have been based on argument structures acquired conservatively from adult speech" (1989, p. 220). However, they emphasized precisely the smaller portion of the double-object uses which did not correspond to the input, thereby rejecting "strict" conservativism in favour of the "weak" version; although not frequently, children did demonstrate to use verbs in the double-object construction that they have not encountered previously in the same construction (in the adult input). When Campbell and Tomasello (2001, p. 259) made an objection to their research, stating that a more beneficial way to explore the role of the input would be to look at lexically specific relationships (rather than combine all verbs), they opted for a more sensitive analysis, observing each verb used in double-object datives and prepositional datives, comparing child and adult frequencies for each of the verbs separately. Their analysis revealed that children and adults had the same preference for 21 verbs (out of the total 26 they observed) when it comes to the choice of alternation in which they tend to use it (whether it is double-object dative or the prepositional dative). Their preferred choices in ditransitive alternations differed with 5 verbs, 3 of which the child preferred to use in the prepositional dative, whereas the parent did not (for the other 2, it was the other way around). Finally, their conclusion was that children indeed tend to learn the double-object dative first because this was what they heard in the input most often (also, the input that they received from their parents seemed to favour the double-object construction in the ration of 2 to 1).

When it comes to previous research on verbs that appear in ditransitive constructions, such as *give*, *bring*, and *show*, Bowerman (1990) looked into longitudinal data on the spontaneous speech produced by her two English speaking daughters in order to see whether there are any innate linking rules that might facilitate the acquisitional process. Her results supported the claim that linking rules were learned rather than inherited innately; when it comes to mapping thematic roles onto syntactic functions, the children had difficulties with verbs that one would presume were easier to link than some others. Instead, she proposed that children clearly learn linking rules and that their early production of errors ought to be interpreted as a side-effect of "overregularizations of a statistically predominant linking pattern to which they had become sensitive through linguistic experience" (ibid.).[28] Overall, Bowerman's diary-based

[28] The proposition that linking rules might be innate was famously suggested by Pinker (1984; later revised in 1996), who hypothesized that children somehow seem to have a pre-set understanding of lexical and syntactic categories such as N, V, NP and VP), whereby they are able to recognize the major thematic roles in the sentence. According to this account of grammar acquisition, the child's first challenge in the learning process is to locate

data has revealed to her that canonical verbs do not present a greater challenge for children than non-canonical ones in terms of linking; for instance, as soon as one of her children (Christy) was able to master the construction with one verb and two arguments, she had no problem to manage a variety of verbs just as well. When it comes to ditransitives, one child (Eva) preferred to use the prepositional dative, she acquired both patterns rather simultaneously, while the other (Christy) favoured the double-object construction. She interpreted the results of her research as evidence for the fact that linking rules are not innate, adding that the statement of language learners being guided by innate linking rules would not be suitable for interpreting the data which suggest that children seem to have more problems with mapping thematic roles onto arguments which occur most frequently in the input (and are thus prototypical in a sense). Furthermore, Bowerman concluded that the very idea that linking rules need to be inherited or learned is superfluous in the context of mastering certain constructions, i.e. a child could become a fluent speaker without necessarily grasping the 'agent-subject' relationships (ibid.). Thus, it might be more beneficial to view these abstractions as something that children end up with, rather than something that they start out with (Brown, 1973, p. 122).

Some of the research focused on the processing difficulty in ditransitive constructions when it comes to their information structure and its interdependence with syntax, such as the effects of introducing the new information in prepositional datives and double-object datives and examining the degrees to which the processing difficulty varies (see Brown et al, 2012, p. 199). The data consistently showed that the information structure tends to be directly associated with the syntactic arrangement (see Clifton and Frazier, 2004; Brown et al, 2012; Hilpert, 2020). As can be expected, the findings based on the reading time of altering ditransitive constructions suggested that it is more difficult to process ditransitive constructions deviating from a conventional pattern of introducing new information.

The studies dealing with the processing difficulty in terms of distinct linguistic patterns, at least when it comes to double-object datives and prepositional datives, date back all the way to 1970s and early 80s (see Fischer, 1971, Osgood and Zehler, 1981, Roeper et al., 1981), with

these syntactic functions and categories in the input (despite already possessing knowledge on them), and in order to do so, it is presumed that they rely on the innate associations between semantics and syntax (the inferential mechanism is also referred to as ‘semantic cueing’ and it happens at the very outset of language acquisition). Given that they possess an innate knowledge when it comes to ‘linking’, children figure out the positioning of subjects because they know which constituents to look at when searching for thematic roles such as that of the ‘agent’, i.e. they know subjects are most likely to be agents (see ‘semantic bootstrapping’ in Pinker, 1994, 1996).

most of them investigating the children's ability to produce and imitate different types of datives with full NPs as their objects. The results tend to suggest that it is more difficult to process double-object datives with full NPs at their objects than the prepositional alternation. This, in turn may play a significant role in the interpretation of the acquisitional trajectory children display. Gropen et al. (1989) noted that children produced some prepositional datives with greater ease than double-object datives, but such conclusions should be restricted only to cases where datives take full NPs as their objects (e.g. *The giraffe sent the hippopotamus the elephant* vs. *The giraffe sent the elephant to the hippopotamus*), and this need not necessarily imply that double-object datives are acquired later than prepositional datives.

In the context of this research, these pioneer studies are important because they opened up a discussion not just on the order of acquisition, but also on the diversity and density of early speech, as well as questions pertaining to restrictiveness, abstractness/concreteness and context-relatedness of the earliest usage of verbs and other lexical items. Although the data used in this research is not sensitive enough to explore the small feats in the routes of acquisition, the sheer size of the corpora should yield interesting new results on the lexical enrichment in particular, but also put into perspective the differences in the syntactic complexity across and within the observed age groups.

An important research that this thesis also builds upon is the work by Goldberg, Casenhiser and Sethuraman (2004), who compared the adult speech and children's early speech using the data from the CHILDES corpus. Challenging the idea that the input is not sufficient for children to deduce the required representations from it, they analysed the proportions of verbs in different constructions, including the ditransitive one. What they found was that the verb *give* was by far the most frequent in children's speech, mirroring the speech of adults. They argued against the interpretation that the sole reason for this is the fact that this is the way *give* is distributed across ditransitives in language in general, and advocated for the relevance that input has on governing the acquisition of language, or more precisely, development of the ditransitive construction in children's speech. They formulated a hypothesis, which they later additionally confirmed in an experiment, that the occurrence of particular verbs in particular constructions that are highly frequent in the input allows them to discern a correlation between the meanings of those verbs in their constructional patterns and the patterns themselves, "giving rise to an association between meaning and form" (Golberg et al., 2004, p. 299). They further suggested that this might serve as evidence for pathbreaking general-purpose verbs that later

facilitate and guide the acquisitional process, as previously hypothesized by Ninio (1999), although their take on the matter differs considerably from his. They highlighted that they do not necessarily consider that these general-purpose verbs need to be acquired before the others in the developmental trajectory, nor that it is important that these verbs have such general meaning, adding that their research considers the sole effect of frequency - "the existence of a high-token-frequency exemplar facilitates the acquisition of constructional meaning" (Golberg et al., 2004, p. 307). It just so happens that these general-purpose tend to be those which are present in the input most precisely because their meaning tends to be widely applicable and highly relevant. Building upon this study, Casenhiser and Goldberg (2005) conducted experiments that tested for the speed of mapping between meaning and novel forms. Their research revealed that children (as well as adults) needed only three minutes of training in order to retrieve the novel meaning and its association with the newly introduced pattern. The research also showed that the input where one verbal token accounted for the balance of utterances was especially facilitative for the acquisition process, i.e. mapping between meaning and novel forms. The results of the experiment point towards the cognitivist takes on the process of acquisition, which attribute this effortless mapping precisely to general cognition. However, the fact that so little input can result is such rapid mapping may be perceived as counterintuitive to usage-based accounts which insist on the density of input as one of the main predictors for early language development. Here, Goldberg (2006, p. 90) takes a clear stance that rapid mapping with so little input needs not necessarily collide with usage-based approaches to language acquisition and evidence of the so-called conservative learning.

3.2. Ditransitive constructions – results

3.2.1. Extraction method and the examples of ditransitive constructions

The major part of the corpus research consisted in locating the constructions within the corpus, extracting the instances located, and finally, filtering out the examples whose structure indeed represents the target construction (the majority of instances only superficially resemble the pattern, but reveal themselves as false positives after further inspection). The crucial tool in this extraction process is the POS (Part-Of-Speech) tagging provided by the platform Sketch Engine. As opposed to periphrastic causatives and relatives, ditransitive constructions are less anchored to specific verbs or items, and more to the general syntactic pattern that typically

entails a transfer of sorts; in this case, the double-object ditransitive *NP verb NP NP*. The verb list is not necessarily fixed (or as restricted) as in causatives, where only a handful of English verbs denote the periphrastic variant. Thus, the location and extraction of the construction from the corpora was tailored accordingly, i.e. using the POS tagging tool, the corpora was searched for instances that include various combinations and types of NPs, such as those that contain determiners and modifiers, as well as the standalone nouns or pronouns (see Appendix 10.1 for POS syntax).

The initial step in the research (as already stated in Section 2.3.1) was to filter out the targeted examples from the examples found in the corpora that accidentally happened to overlap in their syntactic patterns. Table 2 illustrates the final product of this process and contains some the examples selected from the corpora that represent the utterances produced by children and ultimately included in the analysis. Just from the superficial look at the examples, it becomes clear that these are far from the textbook examples of ditransitive constructions, frequently diverging in terms of grammatical accuracy, often containing unnecessary repetitions or omissions of particular words. Naturally, the spoken language significantly differs from the written one, with the former often relying on the context, situational character and cues in order to convey the message at hand. However, despite the differences, many of these, when stripped of redundancies in their surrounding elements, still represent what you might consider the canonical, if not prototypical, ditransitive constructions.

Table 2. Examples of (candidate) ditransitive constructions found in the CHILDES corpora

Corpus	**Age (years; months; days)**	**Produced ditransitives**
BROWN	2;1;0	*But I go write you a lady now.**
MANCHESTER	2;1;21	*Bring lawnmower a walk.*
SUPPES	2;10;21	*Who gave me this little doggy?*
MANCHESTER	2;5;20	*I gonna get him the cold water.*
MANCHESTER	2;6;19	*Make tea the baby too.**
MANCHESTER	2;6;5	*And bring me a pig with it.*
BLOOM	2;8;12	*Make you car make yours.*
GLEASON	3;0;1	*A but Mom would you give me some more milk?*
WELLS	3;3;2	*I tell you a story about this.*
BROWN	3;3;4	*Can I put it in here and make funny man foot?***
WEIST	3;3;6	*Do you show me something when I was a baby again?*
BELFAST	3;6;21	*Santa's going to brought me them.*

MACWHINNEY	3;7;3	*Daddy I want to tell you another question.*
BROWN	4;2;17	*I got ta show Gil some of my pictures.*
MACWHINNEY	4;2;29	*That's good so we can give Rossy some when he gets home.*
WEISMER	4;6;0	*And I got different monsters legos toys and wow.***
HALL	4;6;0	*And den and den I be a and den when I be a nurse I give people needoos.*
HSLLD	5;10;16	*Mama give us some Coke while you're up.*
MACWHINNEY	6;3;10	*Dad you promised me frozen and I never got it!*

*dubious but retained in the analysis **removed from the final dataset

In some of them, it can be rather unclear as to whether they constitute a ditransitive construction, but in the majority of cases, the produced constructions seem clear-cut. For instance, examples such as *But I go write you a lady now* or *Make tea the baby too* (see Table 2) from the Brown or Manchester subcorpora may require further consideration with regards to their intended meaning, and consequently, their classification as ditransitives. In the latter case, although the child confuses the order of the theme and the recipient, we may still insist that this constitutes a valid effort on behalf of the child to produce a ditransitive construction, and for this purpose, it is left as a part of the data to be used in subsequent analysis. In the case of *But I go write you a lady now*, the researcher is faced with a more daunting task when it comes to assigning this particular utterance a corresponding construction. There are several possible interpretations of this particular sentence that appear to be most plausible; (1) the recipient is represented by the pronoun *you* and the child has some theme in mind that they (probably) wrongly refer to as *lady*; (2) the recipients are both the pronoun *you* and the *lady*, who represent the same person, while the theme has not been vocalized by the child; (3) the recipient is represented by the pronoun *you* and the rest of the sentence has nothing to do with the act of writing, but represents an interruption in the child's train of thought. The point, however, remains that the uttered construction invites a transfer of sorts, and includes two overt objects according to at least one of the possible interpretations. For this reason, the example was also included in the overall analysis, although in the great scheme of things, when analysing the overall frequencies of particular constructions, the inclusion or exclusion of examples problematic for the interpretation does not represent an issue or significantly changes the totality of data. On the other hand, sentences such as *Can I put it in here and make funny man foot* and *And I got different monsters legos toys and wow* represent examples which caused certain dilemmas during the process of filtering out the false positives, and were ultimately eliminated from the final dataset. The former seems to be a case of compounding, rather than

saying that the *funny man* was indeed made a *foot* by the child, whereas the latter seems to be an example of listing things instead of implying that *legos toys* were somehow transferred to *monsters*.

3.2.2. Structural complexity and similarity

When targeting the double-object ditransitive constructions in the analysis, the extraction of data mostly revolved around different patterns that follow ditransitive verbs, that is, the heads of noun phrases that occupy the object positions and the modifying elements that precede them. In this sense, Table 3 represents the initial step in describing ditransitive constructions found across the three age groups, but more importantly, it provides a synthesis of the data pertaining to syntactic (and lexical) complexity of the ditransitive constructions observed across the three studies age groups. Although the complexity in this case is to be attributed more to the nominal elements, rather than the construction per se, the idea behind the outlined data is to provide some context on the complexity of both nominal elements in object positions, and consequently ditransitive construction as a whole. It covers a wide range of patterns, such as patterns including sole nominals or pronouns, modified nominal, preceded by adjectives or determiners, and even clauses at times (which frequently occur with the verbs *tell* and *show*). For instance, the pattern most frequently occurring in the three age groups - V-PRO-MOD-N, may stand for sentences such as *He gave him a present* or *He told him that he should give him a present*. Other patterns such as V-PRO-N can be represented by utterances such as *He gave James present*, V-N-MOD-N by *He gave James a (nice) present*, V-PRO-PRO by *He gave him it* etc. (for some of the corresponding examples observed in the CHILDES corpora, see Table 4).

Table 3. Distribution of different ditransitive patterns across age groups (relative frequencies normalised per million tokens and rounded to the closest unit)[29]

	0-3	**4-6**	**18+**
V-PRO-MOD-N	333	627	1068
V-PRO-N	90	197	135
V-N-MOD-N	28	43	133
V-PRO-PRO	62	90	48
V-MOD-N-MOD-N	8	30	38
V-N-N	22	18	21
V-MOD-N-N	2	5	7
V-N-PRO	4	5	5
V-MOD-N-PRO	0	2	0
TOTAL	549	1017	1455

Table 4. Examples of the produced ditransitive constructions in the CHILDES corpora (patterns from Table 3 indicated in square brackets)

Construction pattern	**Example**	**Corpus**	**Age**
V-PRO-MOD-N	*My mommy [gave me the candles from Charlotte] cause those candles were going down all was going to melt.*	MACWHINNEY	3;3;15
V-PRO-N	*Okay then you'll have to [earn me money].*	KUCZAJ	4;9;12
V-N-MOD-N	*I got ta [show Gil some of my pictures].*	BROWN	4;2;17
V-PRO-PRO	*Will you [read me it]?*	MACWHINNEY	3;8;3
V-MOD-N-MOD-N	*And a and a little woman [giving a little child a mouse].*	MACBATES	5;0
V-N-N	*Hulk's girl friend we [call Hulk Woman].*	MACWHINNEY	3;10
V-MOD-N-N	*I [give big tiger a bracelet].*	MACWHINNEY	2;6;17
V-N-PRO	*You [give mummy one]?*	BELFAST	18+
V-MOD-N-PRO	*Daddy [give the ball me].*	MANCHESTER	2;1;7

Furthermore, as expected given the initial outline of the specific pattern frequencies, the Spearman's rank correlation reveals very high values for inter-group similarities (N=9). The value is slightly higher for the similarity between 4-6 age group and adults, though the margin

[29] The data on frequencies concerning the distribution of ditransitive construction patterns across age groups has already been published in Proroković and Malenica (2023). Also, note that the zeros in the table do not necessarily represent absolute zeros (i.e. a complete absence of the pattern), but they do indicate that the pattern has been observed less than 0.5 times across million tokens. The same is true for other tables concerning relative frequencies of different construction patterns in the thesis.

is small (see Table 5). The highest similarity in terms of rank ordering is between 0-3 and 4-6 age groups, but given the very high values for all, the overall take away is that all three age groups exhibit more or less the same hierarchy of use in terms of specific ditransitive patterns with respect to object elaborateness.

Table 5. Spearman's rank correlation of the observed ditransitive patterns across age groups

	0-3	**4-6**	**18+**
0-3	**1**	**0.983****	**0.950****
4-6		**1**	**0.967****
18+			**1**

**ρ=0.683 p=0.05 **ρ=0.833 p=0.01*

The analysis of ditransitive constructions across age groups reveals similar tendencies in terms of their argument structure and complexity. To begin with, the most frequent ditransitive pattern is the one with the pronoun as its first object (indirect one) and the complex nominal as its second (direct) object. Considering that the recipient in the ditransitive construction tends to imply an animate being, it is not surprising that the most frequent selection for the argument takes on the form of a pronoun. Even in the non-prototypical cases of ditransitives with proverbial character, it is not rare to find the neuter, third-person singular pronoun *it* (e.g. *(…) so I feed it this giant piece of mea*t; HALL, 4;6), and sometimes this may be the case due to inverted word order (e.g. *Miriam gave it me*; SUPPES, 2;0;24). When it comes to the theme, the preferred choice in the most frequent pattern is the complex nominal.[30] There is no doubt that even the youngest age groups have mastered the use of all double-object ditransitive subtypes, at least constituency-wise, and that the trends are remarkably similar between the age groups. However, there are just enough differences that would account for the expectations of children's language being structurally simpler, while at the same time allowing us to maintain the position that children's language relies on the input to a degree where it becomes almost imitative in character.

The increase in the overall complexity (see Section 1.4 for a construction-specific use of the term) is observable as the age progresses. The increase can be observed in the most frequent ditransitive pattern V-PRO-MOD-N (e.g. *(…) he's the one who buyed me cotton candy*;

[30] Note that a complex nominal does not necessarily denote a huge degree of complexity; in fact, the most frequent variant is a single noun simply preceded by an article (here designated as MOD, which includes both determiners and adjectives in this case).

MACWHINNEY, 4;4;1), but also in other variations—when the constructions containing 'complex nominals' are grouped together and compared the spike in the complexity becomes even more evident. In the 0-3 age group, the percentage of constructions which incorporated 'complex nominals' in some of the slots equals to 67.46% of the overall uses, in the 4-6 age group it amounts to 69.49%, and in the adult corpora it takes up 85.7%. As can be expected, children prefer to use shorter and simpler constructions, at least in terms of constituents that make up their phrases. In many cases, the variation may be explained simply by the fact that children are more likely to drop determiners before nouns, as well as the fact that adults use more clauses with verbs such as *tell* and *show*.

Furthermore, one needs to be careful when ascribing the syntactic complexity to the construction as a whole. In other words, it is possible to suggest that the measurement of complexity applies solely to the nominal elements included in the observed pattern, and not to the construction itself. Children may very well acquire the construction and only later progress in terms of more complex nominal elements (arguments) that they use as a part of the ditransitive construction. And yet, given that the ditransitive construction tends to be affected by the length of the argument, or at least the choice to realize either the double-object or the prepositional dative construction (see Bresnan et al., 2007), we may safely assume that the complexity of arguments associates closely with the structural complexity of the overall ditransitive construction. Naturally, the data from Table 3 indicates that the complexity of the nominal elements in the ditransitive construction increases with the progression of age. For example, the results indicate that the position of the second (direct) object in the ditransitive constructions tends to be occupied most by a complex nominal in the adult speech, regardless of the fact that it is child directed.

The first look at the data indicates a rise in the total use of ditransitive constructions. Clearly, the amount of overall use of ditransitive constructions incrementally rises with age – first, it almost doubles, and between the 4-6 age group and the adult group it increases by approximately 30%. Ditransitive constructions are evidently mastered by children younger than 4, but the overall use is still significantly smaller than the one in older age groups. What does this tell us about the use of ditransitives and the children's earliest utterances? Given that we are looking at relative frequencies (normalised per 1 million tokens), the data illustrates the fact that the overall language use by either age groups is occupied with a particular portion of produced ditransitive instances. In other words, the more ditransitives there are, the less of other

types of utterances are left in the overall language uttered by the speaker. This in turn means that this type of data reveals also something about what is not included in this analysis, and yet was produced by the speakers. It could be posited that such discordance in the data can be explained by the fact that earliest age groups prefer (or require) to use other types of constructions in their everyday communication. A more accurate assumption would probably be that the overall speech of children recorded in various CHILDES subcorpora contains a lot of incoherent children's talk, such as single, double, or triple-word utterances with little or no meaning.

The distinction between the last two assumptions is particularly important because it leads to two different conclusions. For instance, the first assumption (that children may simply be using more of other types of constructions which then occupy a great part of their overall language production) does not really help in terms of inferring about the overall language complexity of early age groups; this is a part of the reason why this thesis observes the use of other constructions as well. The second assumption (that there is a lot of incoherent children's talk occupying the rest of the produced language), which is probably more accurate, helps put the data into perspective and shows the degree of overall linguistic prowess that we ascribe to particular age groups. Clearly, there is nothing groundbreaking about finding that the language of children younger than 4 might be simpler than that of adults, or that of 4-6 age group. However, it does reveal to what degree this might be the case, and when it comes to ditransitive constructions, it appears that the use of them increases dramatically and progressively between the studied age groups.

Table 6. Distribution of different ditransitive patterns across age groups (relative frequencies normalised per million tokens; some patterns are merged together)

Normalised frequencies per million tokens	**0-3**	**4-6**	**18+**
V- PRO- N/MOD-N	423	824	1203
V-N/MOD-N-N/MOD-N	60	96	199
V-PRO-PRO	62	90	48
V-N/MOD-N-PRO	4	7	5

The data from children's speech agrees with arguments about pronominality[31] as being one of the factors for predicting the realization of the ditransitive (see Collins, 1995, Bresnan &

[31] The term is used to dissociate phrases headed by pronouns of any kind from those headed by nouns.

Nikitina, 2003, Bresnan et al., 2007) that is, the double-object ditransitive is inclined to incorporate the pronoun in the position of the first (indirect) object, but prefers to avoid the pronouns in the final position. When the production rates of ditransitive construction patterns are merged together to better reflect the disparity in the use of (modified) nominals and pronouns with respect to the object positions they occupy (see Table 6), the Chi-square test shows that the difference in ditransitive production rates is significant, $\chi^2 (6, N = 4) = 62.9231$, $p < 0.05$. When the patterns are merged into two groups depending on the part-of-speech of the second object (V-NP-N/MOD-N and V-NP-PRO), the Chi-square test again shows that the difference in ditransitive production rates is significant ($\chi 2=55.32$, df=2, $p<0.01$), with the ratios between the two construction patterns increasing significantly towards the older age groups (this is especially visible in the adult group, where the pattern which takes a noun as its second object is represented to the greatest extent). This is why it is particularly interesting to observe the difference between the age groups in the pattern V-PRO-PRO, which is also suggested to be a non-canonical ditransitive that sounds awkward (cf. Haspelmath, 2015, p. 10); in relative terms, the overall proportion of said pattern is highest in the 0-3 age group. When it comes to 0-3 age group, it takes up around 11% of their overall ditransitive use, in 4-6 it takes up around 8%, while adults only use it in 3% of the cases. One obvious possibility might be related with the lexical richness in the early age groups, where it may have happened that the child could not retrieve the adequate lexeme needed in the situation at hand, and opted for the pronoun instead; instead of saying *give me the teddy*, the child may have said *give me it/him* (e.g. *I want you to give me it*; BELFAST, 2;6;16). However, this explanation is not necessarily sufficient considering the number of arguments even the youngest age groups have been observed to use in the ditransitive construction (see Proroković & Malenica, 2023). An important point regarding the corrective function of the frequency itself, suggested within the usage-based tradition (Lieven, 2010), might be further challenged here, i.e. if we imagine the construction to be a non-canonical one which needs grammatical rectification,[32] we have to think about the reasons which moderate this drop in the overall portion of such construction

[32] The fact that something is rather unorthodox in grammatical terms is not integral for the claim about the corrective purpose of the input, and in fact, it even reduces the strength of the argument to a degree. In other words, if the pattern is a non-canonical one grammatically speaking (e.g. there may be more "elegant" syntactic alternatives for the same expression), it would require an even greater reinforcement in the input in order to be produced by children. The situation in this case is the opposite one – the supposedly non-canonical pattern which is seldom found in the input is produced more in children's speech in relative terms. This can also indicate that the pattern is not so syntactically unorthodox, or that the noun/pronoun distinction is not that relevant in the early developmental context.

use. The reason may very well be that the sole frequency in the input bears the corrective function; low frequency in the input may account for the progressive drop in the use of such ditransitive pattern – the children are able to observe that such use of the ditransitive construction is not common and infer that the pattern V-PRO-PRO is generally best avoided. However, the results still remain counterintuitive, given the greater relative production of the pattern in the younger age groups. It is important to note that a moderate anti-nominal inclination has also been observed in some periphrastic causative construction patterns (see Section 4.2.2), though never to the degree where the use of pronouns overshadows the use of nominals in the younger age groups – especially if such relation is not attested in the input.

3.2.3. Lexical arrangement and overlap

The following part of the analysis explored the most frequent verb choices within the ditransitive construction (see Table 7). As heads of phrases, verbs have been traditionally interpreted as focal points in the acquisition of language and constructions (see Goldberg, 2006). Both ideas of the so-called 'pathbreaking verbs' and 'verb-islands' recognise the importance of prototypical items frequently found in certain constructions, which are acquired first (and if not first, they show a tendency to appear early on) and which tend to have a stimulating role in terms of subsequent language development. The 'pathbreaking verbs' hypothesis (see Ninio 1999, 2003) is, in some respect, similar to Tomasello's early 'verb-island' hypothesis (1992), later rephrased and altered into 'item-based' hypothesis. The two hypotheses differ with regards to whether the relationship between different verbs is fluid within and across constructions; in some sense, the verb island hypothesis views the syntax of a particular construction as initially learned for each verb separately, meaning that the language of children consists of item-specific patterns initially treated separately throughout grammatical development, and only later generalised and understood in terms of constructions. While the 'verb island hypothesis' agrees with rote learning principles, the 'pathbreaking verb' hypothesis resonates more with what is known as the 'Subset Principle' (see MacWhinney, 2005, p. 59) where the grammatical forms are envisioned as parts of the slightly more complex grammars, and where the child generalises the use of a particular verb across competing constructions (e.g. across alternating dative constructions) if the same has been attested in the input (see Fodor & Crain, 1987). However, the two theories agree on the fact that the learning of a particular

construction begins with prototypical verbs heard frequently in the input, which later guide and facilitate the acquisition process of other lexical variations within and across constructions. These results do not directly test the 'pathbreaking verbs' hypothesis, but they certainly show an incredible disparity in the use of certain verbs in ditransitive constructions. The data further corroborates the claim of the verb *give* being the most representative example in the ditransitive construction, and probably the 'pathbreaking' one in most of the cases. More importantly, the data also shows an important role of the following 4 most frequent verbs in the production of ditransitive construction.

This type of comparative analysis provides valuable data relating to the item-based language learning hypothesis (Tomasello, 2000b, 2001c; MacWhinney, 2014), and contributes additional evidence on the relationship between children's produced utterances (output) and the language they heard in their environment (input).[33] In terms of synchronic comparative data, the claim transposes to expectation that certain verbs will cover most of the ditransitive uses, as well as that overall use of ditransitive constructions might be headed by fewer number of verbs in early speech.

Table 7. Distribution of ditransitive verbs across age groups[34]

Verbs/ Age	0-3		4-6		18+	
give	1073	42.65%	798	41.85%	11162	37.25%
tell	233	9.26%	367	19.24%	6295	21.01%
get	374	14.86%	207	10.85%	2654	8.86%
show	157	6.24%	124	6.50%	1885	6.29%
make	186	7.39%	71	3.72%	1405	4.69%
buy	91	3.62%	92	4.82%	950	3.17%
bring	75	2.98%	40	2.10%	867	2.89%
read	68	2.70%	37	1.94%	600	2.00%
find	22	0.87%	5	0.26%	407	1.36%
ask	5	0.20%	20	1.05%	385	1.28%

[33] Tomasello (2001c) himself stated that these claims are best challenged by negative evidence, that is, by observing which items do not appear in certain patterns, despite their otherwise common presence in the input. This type of analysis unfortunately falls beyond the scope of this research as well. However, while it is valid to maintain the usefulness of such data, the method would not be without clear limitations and drawbacks. In the context of CHILDES, considering that we do not possess data on the entire linguistic input that the child has received throughout their lifetime, we cannot derive conclusions, not with absolute certainty at least, about their atypical overgeneralisations or supposedly creative language uses in terms of deviations from what they have heard. It would surely be possible to make a leap of faith based on some justified assumptions, but there is no doubt that it presents methodological issues from a rigid perspective on what qualifies for data comprehensive enough for such testing.

[34] The table on frequencies concerning the distribution of ditransitive verbs across age groups has already been published in Proroković and Malenica (2023).

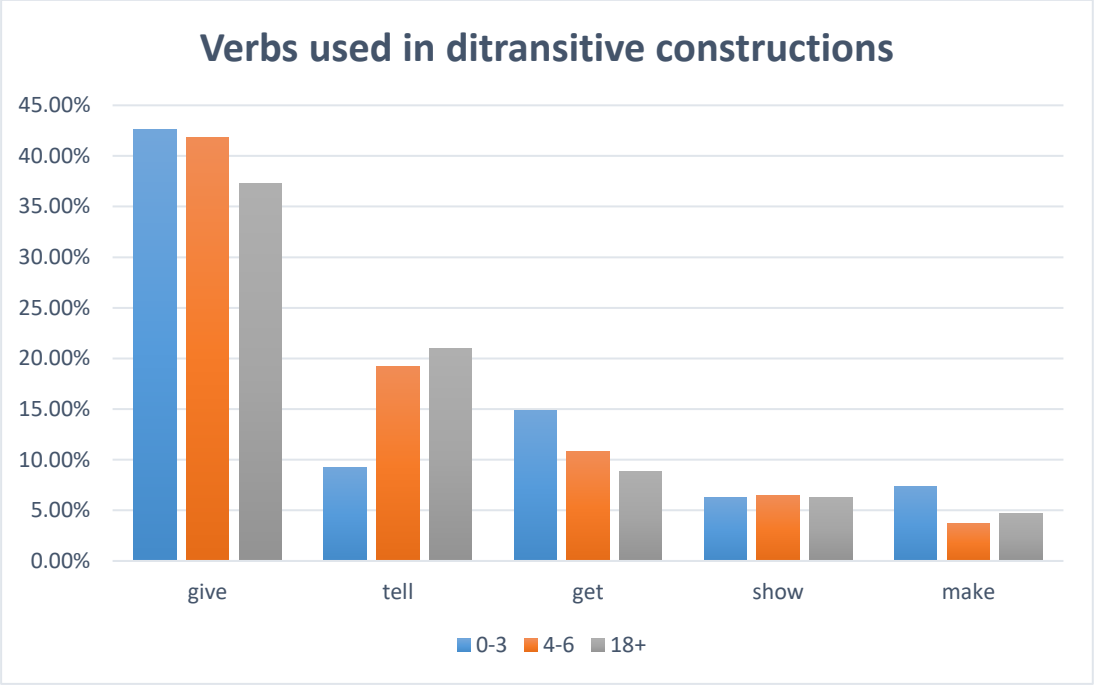
Again, the overall resemblance in the use of particular verbs in ditransitive constructions was evident. By far, *give* is found to be the most frequently used verb in double-object ditransitives, accounting for more than a third of ditransitive constructions in each of the studied age groups. Frequency-wise, the verb is followed by *tell, get, show, make* and other, with relatively small discrepancies in terms of their frequency-based hierarchy of usage. The frequency order by which these verbs are used in ditransitive constructions is remarkably similar between age groups. What is more, on a broader scale, the results reflect the exact usage-based expectations of how output would resemble the input: (1) verbs which are most frequent in the adult language should also be the ones most frequent in children's language and (2) despite overall similarity in distribution, children's early language should still focus around fewer verbs than adult language. For instance, if the verb *give* is the most present one in the input, the same should be reflected in the children's output when it comes to ditransitive constructions; furthermore, if we assume that the verbs most frequent in the input tend to influence and guide the acquisitional process of the constructions in which they appear, we should also expect its representation to be slightly more frequent in the early language. This is exactly what the data shows when it comes to the verb *give*. The situation is the same when it comes to verbs *get* and *make,* but slightly different when it comes to *tell* and *show*. This is to be expected, partly because they are not the most frequent ones and the aforementioned reasoning tends to weaken the further you go down the input frequency lane, but the major reasons might be the fact that *tell* and *show* do not constitute a canonical ditransitive construction. In other words, constructions containing these verbs have been described as ditransitives implying a 'mental', rather than physical transfer (see Malchukov et al., 2010).

Furthermore, if usage-based assumptions were incorrect, one could expect at least some lexical differences across age groups within constructions. For example, why would children so frequently choose the verb *give* over the verb *get* in ditransitive constructions. These two verbs share many characteristics in terms of meaning and are frequently used in the same sense. What is more, *give* tends to have a very restricted meaning, whereas *get* appears more flexible meaning-wise, implying a wider range of situations aside giving, including the grammaticalized ones (such as the periphrastic causative one; see Section 4.1.2). If one were to completely ignore the effects of the input, one should expect more competition between the two. Indeed, the *get-give* ratio is larger (the difference in frequency is smaller despite seemingly corresponding

disparity in overall proportion between the two) in the 0-3 age group, but it does not seem nearly enough for us to infer any sort of competition between the two. If anything, it seems to be explained by the fact that adults are simply more lexically diverse in their speech.

If the five most frequent verbs (see Figure 2) in ditransitive constructions are added together, they constitute approximately 80% of the total use across all three age groups; more precisely, the verbs *give*, *tell*, *get*, *show* and *make*, take up 80.40% in the 0-3 age group, 82.16% in the 4-6 age group and 78.10% in the adult group. In some sense, the distribution of verbs in the produced utterances reflects the Pareto principle, although with slightly different rations; for instance, in 0-3 age group, the first 5 verbs take up approximately 10% of different verbs found in ditransitive construction, but in terms of overall use, they take up approximately 80%. The disparity is even more pronounced when it comes to the 'vital few' and their productive potential, as they take up approximately 5% of ditransitive uses but also constitute close to 80% of the overall use of ditransitives.

Figure 2. Most common verbs used in ditransitive constructions



Perhaps the disparity in the usage of *tell* and *show* seems to be one of the most interesting elements of the data, for they are not more salient in the youngest category as one might expect (if one supposes that the most frequent verbs would also be most salient in the youngest category because the older ones tend to increase the range of verbs used in ditransitive

construction). Tomasello (2000a) has hypothesised that language tends to be imitatively learned first, which may entail simultaneous or subsequent efforts to comprehend the utterance.[35] Here, Tomasello carefully elaborates the concept of 'imitative learning' and adds the element of comprehension, claiming that 'imitative learning' is unjustifiably stigmatised in language acquisition research, and that it needs not be interpreted in the context of rigid repetition of verbatim, but rather elaborated with the role of 'understanding' and the ability to execute 'functionally based distributional analysis'. In other words, children need to reproduce the surface linguistic form so that it reflects its conventional communicative intent, thus exhibiting the fact that they understand its underlying function (see Tomasello 1998, 2009). Seemingly, these results counter the idea that certain linguistic patterns are first learned and understood later, given that the frequency of the two verbs in the 0-3 age group clearly does not follow the same frequency-based order of usage. If such hypothesis were true, we might expect that the order will be the same, regardless of the 'understanding' element. But given that understating can play an important role simultaneously to imitation, the results actually confirm some of Tomasello's conjectures regarding early verb usage and language character. The evidence clearly shows that the youngest age group has clearly produced/mastered ditransitive uses of the two verbs, but also that their use is rather restricted when compared to other verbs in the same constructions and the distributions in the other two age groups. The important conclusions that we might draw from this data are the following: (1) The level of abstraction and cognitive capacity surely plays a role in acquisition—in the usage-based canon there traditionally exist an insistence on the mechanisms of reading cues in environment and there is no doubt that more abstract verbs rely on cues less than others, meaning that children should find it harder to acquire such verbs; (2) The two verbs (*tell* and *show*) often imply more complex syntax of the direct object (such as clauses), which has partly skewed the results in favour of adults and (3) Children's language also reflects the variety of needs and situations in which they find

[35] Quote from Tomasello (2000, p. 73): "In either case, the main point is that young children begin by imitatively learning specific pieces of language in order to express their communicative intentions, for example, in holophrases and other fixed expressions. As they attempt to comprehend and reproduce the utterances produced by mature speakers – along with the internal constituents of those utterances – they come to discern certain patterns of language use (including patterns of token and type frequency), and these patterns lead them to construct a number of different kinds of (at first very local) linguistic categories and schemas. As with all kinds of categories and schemas in cognitive development, the conceptual "glue" that holds them together is function…"

themselves, favouring the verbs which require something concrete of their interlocutors (such as *give*, *get* or *make*), thus skewing the results in favour of these verbs.

The similarity that was observed in terms of structural complexity across age groups seems to be even more apparent in terms of lexical choices governing ditransitive constructions. In order to statistically verify these suppositions, Spearman's rank correlation analysis was conducted by taking the 22 most frequent verbs in the ditransitive construction and assigning ranks to them in terms of how frequently they appear in each of the age groups (Table 8).[36] The results of the correlation analysis indicate a strong association between the 0-3 age group and the other two, and a very strong resemblance between the 4-6 age group and the adult group in terms of a frequency-based ranking of ditransitive verb uses. The overall similarity, both between and across all three age groups, is relatively high and statistically significant at the 0.01 level ($p<0.01$).

Table 8. Spearman's rank correlation of the most frequent ditransitive verbs across age groups

	0-3	**4-6**	**18+**
0-3	**1**	**0.697****	**0.707****
4-6		**1**	**0.824****
18+			**1**

**ρ=0.423 $p<0.05$ **ρ=0.53 $p<0.01$*

The fact that the correlation is lower between the 0-3 and the 4-6 age groups than between the rest is also interesting to observe; despite the fact that the two age groups tend to share much of the contextual factors which skew the results in favour of some verbs (e.g. frequency-wise, *feed* was ranked 10th in terms of frequency in both 0-3 and 4-6 age group, while it was only 21st in the group of adults) the correlation is still higher between each of the younger age groups and the group of adults. If we were to paraphrase these results as the competition between contextual requirements and the input in the effect that they exert on the output, the input would be interpreted as the winner. However, this does not necessarily agree with the fact that the use of particular verbs corresponds more between the 4-6 and adult group than between

[36] More verbs could have been included in the analysis but the number of instances after the 22nd verb was too low, especially in the younger age groups, i.e. low to the point where only a couple or so were recorded in their speech, which makes it difficult to distinguish them above chance level accuracy and to rank them accordingly.

the 0-3 group and adults. The item-based character of early language and its reliance on the most frequent verbs in the input should entail that the similarity would be the highest between the youngest age group (0-3) and that of adults, especially given that the correlation analysis only looked at the most frequent verbs in their language. The children are implied to expand the use of their verbs in the constructions they acquire in a stepwise manner, and the frequency of use should reflect these tendencies to some degree—the evolution of grammar might result in giving priority to some other verbs in ditransitive constructions and consequently resulting in greater divergence. Although this initially seems like an argument in favour of the language of the 4-6 age group resembling more that of adults, the analysis only looked at the most frequent verbs, and not all of them. Even if we reduce the analysis to 10 or 5 most frequent verbs in the analysis, because a fewer number may be more appropriate to test the item-based tendencies in early language, the 0-3 age group still differ from the adult language more than the 4-6 does. Even when the verbs implying a mental transfer such as *tell* and *show* are removed, the Spearman's rank correlation analysis still shows a greater similarity between the 4-6 age group and the adults (Table 9).

Table 9. Spearman's rank correlation of the most frequent ditransitive verbs across age groups (without the verbs 'tell' and 'show')

	0-3	**4-6**	**18+**
0-3	**1**	**0.595****	**0.614****
4-6		**1**	**0.759****
18+			**1**

**ρ=0.591 p=0.05 **ρ=0.450 p=0.01*

While the verbs are traditionally viewed as central to constructions (the phrases conform to the properties of head verbs), and to the realization of said constructions in the context of language development and acquisition, nouns tend to occupy a secondary position when it comes to 'item-based' perspectives on the same. While the data suggests that children's initial learning focuses on the words heading the phrases, and that the 'imitative' aspect of language reproduced is best observed on such linguistic items, it also seems worthwhile to explore whether this learning extends beyond said items. In other words, in a construction such as the ditransitive one, we may also look at the similarities between the items that fill up the object slots. It is expected that these noun phrases would be less affected by 'usage-based' side-effects due to pragmatic reasons (different position of age groups in relation to their expressive needs),

as well as the range of possible items that may occupy the object slots in these constructions (see Table 10 and Table 11).

Table 10. Most frequent first (indirect) objects in give-NP-NP

0-3	**4-6**	**18+**
me	*me*	*you*
you/ya	*you/ya*	*me*
him	*him*	*him*
her	*her*	*her*
them/em/dem	*us*	*it*
us	*them/em/dem*	*them*
it	*boy*	*mommy/mummy*
mommy/mummy	*girl*	*us*
baby	*dog*	*baby*
daddy	*it*	*daddy*

In the first object position, the representation of noun phrases is completely predominated by the use of pronouns. It is particularly interesting to observe that the first-person pronoun is most frequently represented in 0-3 and 4-6 age groups, while the second-person pronoun occupies the primary position in adult speech. This is a consequence of two factors: the fact that the adult speech is child-directed in character (hence the move away from oneself), and the fact that children tend to be self-centred in behaviour, and consequently their language production. More importantly, the most frequent items in the 0-3 age group are mostly restricted to recipients in their immediate surroundings. Even the pronoun *them*, which outscores *mommy* and *daddy* as far as frequency is concerned, usually entails a certain level of physical proximity (e.g. *Mummy find them room*; MANCHESTER, 2;11;15). On the other hand, in the 4-6 age group, the children grow more flexible in their use, and start using general nouns which do not necessarily imply a certain degree of proximity, such as *boy*, *girl*, *dog*. The correspondence in the most frequent items is also higher between the 0-3 age group and adults, than between the 4-6 age group and adults.

Table 11. Most frequent second (direct) objects in give-NP-NP

0-3	**4-6**	**18+**
that	*some*	*kiss*
some	*money*	*hug*
it	*one*	*some*

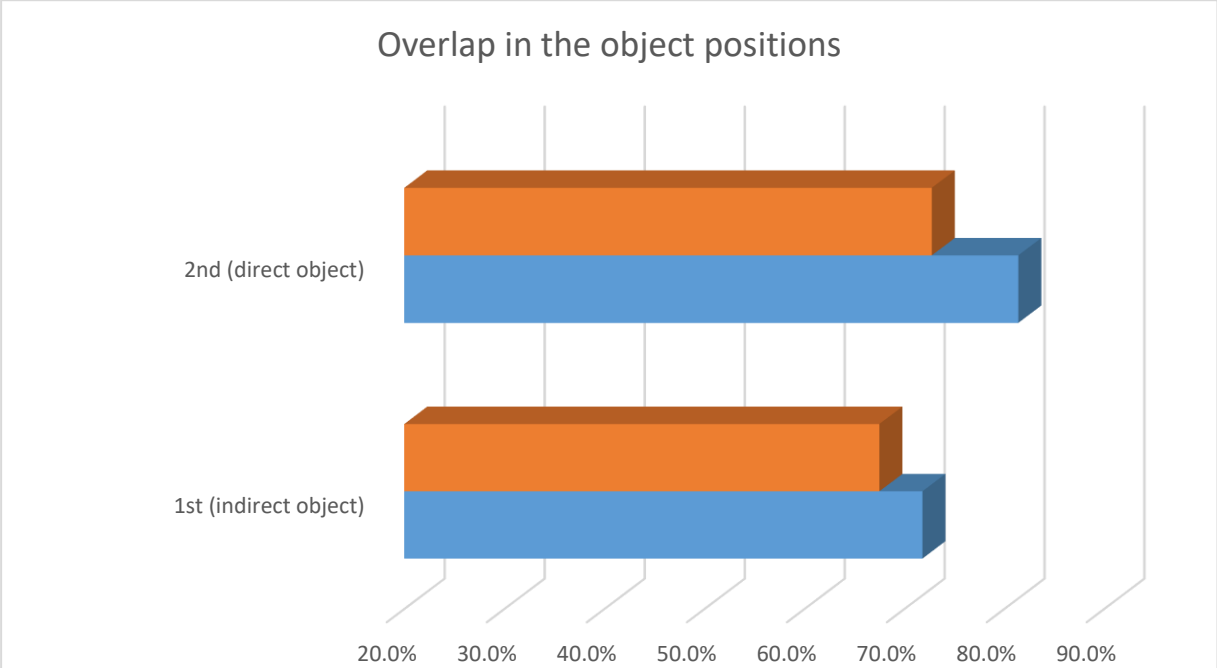
this	*it*	*one*
one	*that*	*money*
hug	*this*	*cuddle*
kiss	*piece*	*something*
bath	*kiss*	*that*
money	*food*	*bath*
present	*something*	*drink*

When it comes to cross-age comparison of second object representation, also Figure 3 shows a certain degree of correspondence across age groups despite the fact that the 'theme' slot is least restricted in the range of items that can occupy its position. Despite a number of items that can stand in the position of theme, we find *some, that, one, kiss* and *money* among the most common ones. While it is understandable that less semantically restricted items would be more frequent as they are able to replace a wider range of nouns, it is indicative that some of the more specific nouns such as *money* and *kiss* are observable across all three age groups among the most frequent ones.

When the entire range of nouns is taken into account, the analysis reveals a greater overlap between the age group 0-3 and the adult group (see Figure 3). In total, out of all indirect objects produced by children, 71.8% have been observed in the input, while the remaining was unique to their production of ditransitives.[37] On the other hand, the numbers are slightly smaller for the 4-6 age group, which, as expected, demonstrates more "independence" with regards to their use of ditransitives. This is even more observable in the second object position, where the difference amounts to almost 9% between 0-3 and 4-6 age group, i.e. the proportion of nouns in the second object position found in the adult input is considerably greater in 0-3 age group than in 4-6 age group.

[37] This does not imply that the children have overgeneralised, or that they have been particularly innovative in their use of the ditransitive construction, since it is impossible to know what the children have heard in their linguistic environment outside of the experimental setting. Still, the data is indicative of (at least) their immediate tendency to reproduce what has been heard.

Figure 3. Proportion of nouns produced by children that were also recorded in adult productions (input)



	1st (indirect object)	2nd (direct object)
4-6 nouns observed in the input	67.5%	72.8%
0-3 nouns observed in the input	71.8%	81.4%

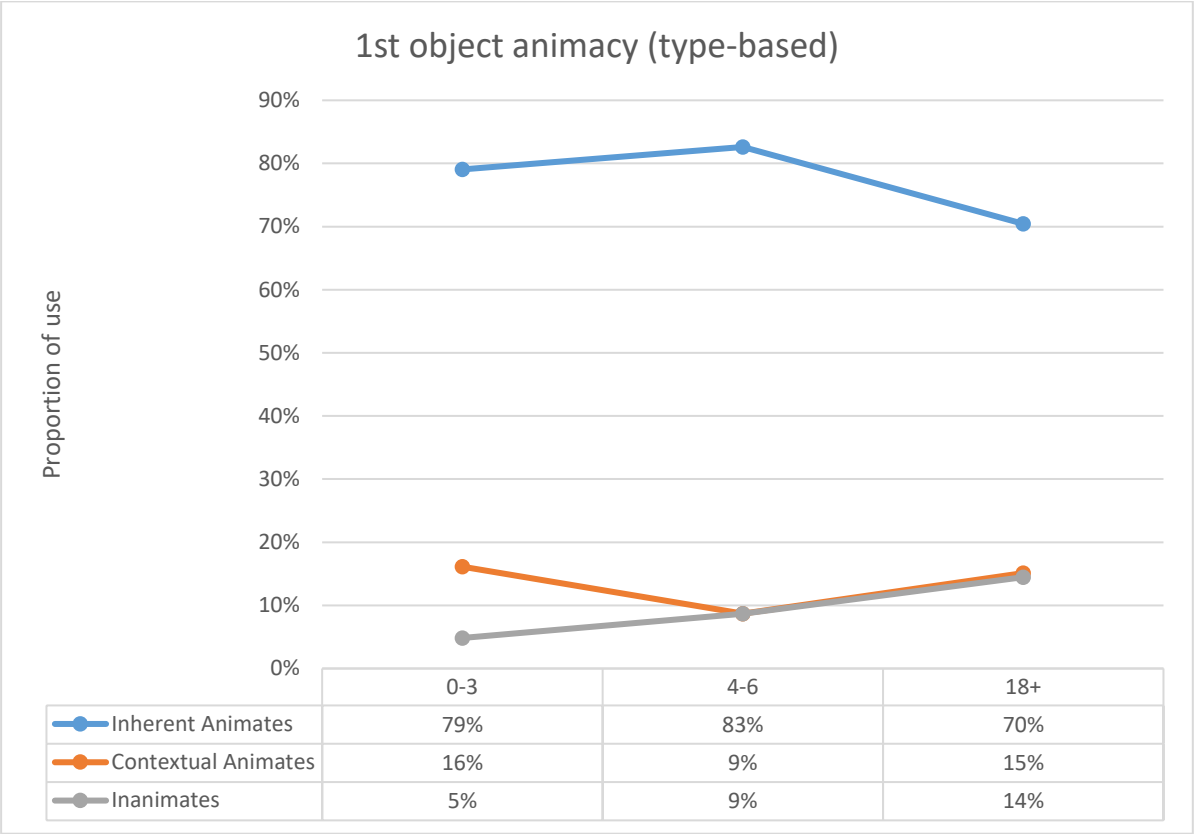
It is also interesting to see that there is a greater overlap in the second object position (theme), which is somewhat counterintuitive given that the slot accommodates a wider range of possible items (e.g. in the 0-3 age group, the number of different types in the first object position is 62, whereas in the second object position 312). One could argue that this may be a side effect of the experimental setting in which the data collection was carried out, but it is not clear why the argument would be any less true for the first object position which represents the recipient. In any case, the data clearly shows a tendency of the youngest age groups to reproduce, if not rely upon, their immediate linguistic environment in producing very specific patterns of language (such as ditransitives) even when it comes to items which are not traditionally seen as central to the construction itself.

3.2.4. Animacy

The double-object ditransitive construction is said to include specific semantic constraints on both first and second object slots. Although one can provide a number of examples that go against the anticipated character of the items that occupy these positions, especially when various idiomatic expressions are taken into consideration, it is safe to assume that the recipient should predominantly be occupied by animates, and that the theme should predominantly be realized through inanimate objects.

The data on first object animacy in the ditransitive pattern *give*-NP-NP confirms the conjectures about the anticipated animacy of the nouns occupying this position in all age groups when lumped together into three categories (inherent animates: common nouns + names; contextual animates: animate toys + body parts; inanimates). First, all age groups share the great majority of inherent animates in the recipient slot. The rest is divided between contextual animates and inanimates, and in a sense – this is where the age groups differ between themselves to some degree (Figure 4). By far, inanimates are least represented in the 0-3 age group, and most represented in the adult group, with the rate of use doubling with the progression of age groups. Inanimates are certainly less represented as recipients in general, and as such depart from common ditransitive constructions (they usually include phrasal expressions such as *give it a try/shot/go*). One line of reasoning would suggest the results are counterintuitive; i.e. children might be more prone to employ inanimate recipients due to overgeneralisations of sort. Still, since there is still a fair amount of expressions that contain some type of inanimate, solely looking at the numbers might not be the best way to confirm such suppositions (even if there are overgeneralisations that do not fit the canonical ditransitive use, the numbers would probably be far too small to reflect greatly on the overall cross-age proportions). Instead, the proportions are best interpreted as a suggestion that adults appear more flexible with the use of inanimates in 1st object position than children, especially the youngest age group.

Figure 4. First (indirect) object animacy



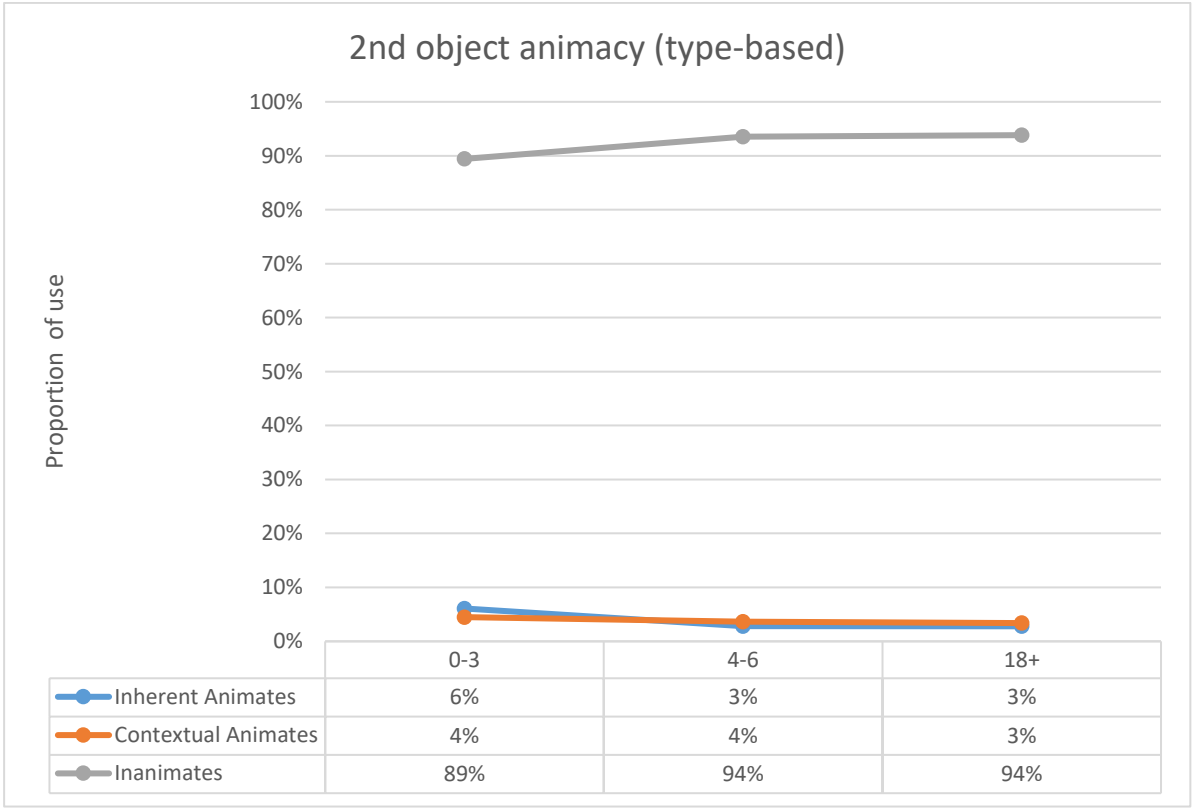
A more detailed look into first object animacy tells us that the recipients are most commonly represented by names (ANp) in 0-3 age group and adults, whereas the most frequent choice in 4-6 age group are general animates (ANc; see Table 12). Animate toys (ANt) occupy the third position in 0-3 age group, while in 4-6 age group and adult group, they share the third spot (in terms of frequency) with inanimates. Body parts (ANb) are almost non-existent in the first object position. Again, 0-3 age group resembles the adult one more in terms of their most frequent choices, as predicted by the first hypothesis, while at the same time they remain reluctant to employ inanimates (INAN) to a greater degree. The results are in line with the usage-based framework, whereby children remain reluctant to produce patterns seldom encountered in their input – or at least – they tend to produce it even more rarely than these would otherwise be encountered in the input. Early language is primarily guided by the most frequent prototypical instances, and as the children grow older, the range of patterns to which their language use opens up expands significantly.

Table 12. First (indirect) object animacy

	0-3	4-6	18+
ANc	39%	54%	24%
ANp	40%	28%	47%
ANt	16%	9%	14%
ANb	0%	0%	1%
INAN	5%	9%	14%

As far as animacy of the second object is concerned in the ditransitive *give*-NP-NP, the cross-age group comparison surprisingly indicates a more constrained use of object types with respect to their animacy. While the recipients were mostly represented by inherent animates, the themes of ditransitive constructions were overwhelmingly represented by inanimates. In the 0-3 age group, they appear to take 89% of overall use, whereas in the 4-6 and adult age groups, they surpass the 90% margin. In contrast to other findings so far, the 0-3 age group does not show the greatest saturation in the most frequent category, but instead demonstrates greater dispersion (albeit to a small degree) between the two remaining categories – inherent and contextual animates.

Figure 5. Second (indirect) object animacy



One possibility is that, prior to 4 years of age, children remain more flexible with second object realizations for some reason; what can be transferred becomes more constrained with the progression of age (which seems opposite to the data on first object animacy). The less conservative use of second object animacy suggests that children possibly go beyond the input and that their perception of what qualifies for a theme in the ditransitive *give*-NP-NP is possibly determined by other factors. To be fair, the resemblance between the two groups is striking, and there is no particular reason to suspect considerable differences in the treatment of items between the three age groups. Nevertheless, the small differences could be triggered by some linguistic, or even extralinguistic factors. For example, it is possible that children early on believe that anything can be given, regardless of whether the object of transfer agrees with the transfer itself; that is, the case where animate themes are employed are relatively rare in English language, and when they are employed – they require some compliance by the person who is being given, or a greater effort (some sort of consideration) at the hand of the person who does the giving, and potentially a greater willingness to receive what is being given at the hand of the recipient (e.g. *I gave my husband a son/a dog*). This usually assumes either shared knowledge between the speakers (cf. Kecskes & Zhang, 2009) or children's ability to read other people's intentions (cf. Tomasello, 2001b), and children's egocentrism and lack of common ground comprehension may result in the effect observed (no matter how small).

3.2.5. Lexical richness

Mean length of utterance[38] and lexical density[39] were calculated in order to provide more information on the structural and lexical complexity of the sentences containing ditransitive constructions. Although the measures are directly related with the ditransitive constructions, they capture the sentences in their entirety (as they are marked in the CHILDES corpora) and many of them contain other types of constructions given that they may consist of more than one clause, or have ditransitive construction embedded in more complex syntactic

[38] Since the intent was to capture all of the sentences used in the ditransitive pattern (as is the case with subsequent analysis of other constructions studied), MLU was calculated with the help of the CLAN program that is not restricted to syllables or characters as most automated text analysis tools available online, but is able to morphologically parse the text once the MOR grammar is integrated and calculate the required measures.

[39] Lexical density is calculated as the ratio between content or lexical words (nouns, verbs, adjectives, adverbs) and the total number of words.

patterns. Additional constraints on the results are imposed due to the fact that the corpora represent transcripts of spoken language, and the choices where to end a particular sentence (to put a full stop) in the process of transcription may not have always been consistent or carefully considered. Nevertheless, this type of information ought to provide context in terms of ditransitive constructions' immediate syntactic and lexical environment.

Table 13. General overview of the sentences containing give-NP-NP across age groups[40]

	0-3	**4-6**	**18+**
Token count	6975	6960	99654
Word type count	682	799	3283
Character count (without spaces)	25215	24960	365272
Utterance count	1014	798	11205
Open class words	3008	2963	43926
Closed class words	3860	3868	53936
MLU (morpheme-based)	7,16	9,21	9,30

The results indicate a rise when it comes to the MLUs of the sentences containing ditransitive constructions with the progression of age. When it comes to the spoken language of children aged from 2;6 to 4;0, the previous data shows an increase in average MLU (calculated in words per sentences) going from 2.9 up to 4.1, and when it comes to children aged 4;0 to 6;11, it progressively increases from 4.1 to 4.7 (see Rice et al., 2010). Naturally, the average MLU in our data is considerably higher (morpheme-based MLU is 7.16 for 0-3 age group, 9.21 for 4-6 age group and 9.3 for the adult group), partly because the limitation on the type of sentences included (those containing ditransitive construction) and partly because of the aforementioned reasons (the occasional problematic identification of actual sentence-endings). Despite the amount of nonce and unintelligible words, which were recorded in CHILDES and undoubtedly more frequent in younger age groups (although CLAN eliminates most of them in the analysis), and which prolong the sentences at least in terms of their length, the results still indicate an increase with the progression of age. In other words, children produce shorter sentences than adults when it comes to those utterances containing ditransitive constructions, and those older than 4 produce longer sentences than the younger ones. The increase is much

[40] Note that the token and word type counts do not include repetitions or revisions (for this table, and the subsequent tables for the corresponding constructions), as set by default in the CLAN program and the applied 'eval' function (for detailed explanation, see MacWhinney, 2023a, pp.132-133) Also, the function excludes gesture-only utterances from the total utterance count, while the morpheme-based MLU evaluation excludes morphemes in utterances representing unintelligible words transcribed as 'xxx', 'yyy' or 'www' codes.

smaller between the spoken language in the 4-6 age group and that of adults, but it is apparent between the two younger age groups observed; when it comes to word-based MLU, the difference almost amounts up to entire two words per sentence. This is a significant increase in the context of the data provided by Rice et al. (2010) which revealed how long it takes for children to increase their sentences by approximately two words, and how this progression somewhat slows down after 4 years of age. The results of this research are consistent with their findings, but in light of additional evidence from the adult group, they also suggest a potential approach to MLU's growth limit; that is, there are three possible explanations: (1) 4-6 age group comes very close to the adult language in terms of how elaborate their spoken utterances tend to be, (2) the increase might be bigger if the adult speech was not mostly child-directed in character, (3) MLU is not a particularly reliable measure of language development. The last argument appears least credible considering the bulk of research that suggests otherwise (see Rondal et al., 1987; Rice et al., 2006; Wieczorek, 2010; Pavelko & Owens, 2019 etc.).

So the question is, where exactly do these sentences differ across age groups? Do children produce shorter sentences because they find it harder to retrieve crucial content words? Is it because their language is poorer in adjectives or adverbs, or is it the function words that they tend to omit? The data on lexical density provides certain clues on where exactly their spoken language appears to differ. The similarity is not to be disregarded considering that lexical density can oscillate substantially, albeit more observably in terms of spoken-written language (Ure, 1971; Halliday, 1985). When it comes to lexical density on its own, some research has shown that it does not tend to be sensitive to age as expected, or that is beneficial to observe it on its own when examining cross-age group differences (Johansson, 2008). However, if this data is considered in the light of MLU, it can indicate where the differences between the studied age groups tend to occur.

In other words, what can we infer from the fact that the utterances between age groups differ in length but do not differ in lexical density? First, it means that the expansion of sentences in the 4-6 age group and adults does not necessarily contribute to the number of content words – given that the ratio stays similar (see Table 14), we could conclude that function and lexical words participate equally in the sentence-expansion process. Moreover, the lack of difference between the observed age groups might suggest that children tend to drop some of the function words from their speech (which would agree with one earlier assumption regarding article dropping); for example, if we imagine the adult language to be saturated more with

modifiers such as adjectives and adverbs, we should also expect a higher level of lexical density in those age groups. Considering that this does not happen, the data might suggest that the reason for this might come at the expense of functional words such as articles, which might be dropped from children speech, and which could in turn inflate the final numbers on lexical density. In fact, this is not surprising at all – despite their significantly poorer vocabulary, children tend to orient themselves on content words in their speech (especially early on), elaborating them with gestures to convey the final message (e.g. *daddy look*, *doggy*). If anything, when "gibberish" is removed from early language, content words would be expected to dominate (especially in the youngest age groups). On the one hand, we have a rather poor lexicon in the youngest age groups which negatively affects lexical density as they find it hard at times to retrieve the content words they require, while on the other hand, we have a tendency of early language to orient itself on the heads of phrases that seem to carry most of the meaning. In the adult language, the situation is somewhat opposite, where the lexical prowess is maximised along with the syntactical one, resulting in sentences which are both well-formed and well-informed. Data on both MLU and lexical density seem to agree well with such view on the opposing processes which participate in determining lexical and structural complexities of adult and early speech.

When it comes to the basic measure of lexical density, there is little or no variation between all age groups in utterances that contain (or centre around) the double-object *give* ditransitive constructions (Table 14). The values seem to be somewhat surprising, given that an isolated ditransitive construction should be more favourable to content words in the LD ratio. For example, a sentence such as *give the man a job* would give an LD of 0.6 (three content words and two grammatical words). In fact, the LD values for isolated *give* ditransitive constructions in 0-3 age group was approximately 0.66. This means that ditransitive give constructions tend to take up only a segment of the utterances produced, and that the rest of the utterance compensates for lack of function words. What is more, we could expect lexical densities in this case to be even higher in the younger age groups because they should produce shorter sentences without going beyond the parameters of the construction at hand. In that sense, a greater number of such short but concentrated utterances would produce higher values of LD.

Table 14. Lexical richness for the ditransitive give-NP-NP

	Lexical density and sophistication			Lexical diversity: range and variation				
	LD	LS1	CVS1	NDW	NDW-ER50	MSTTR-50	RTTR	SVV1
0-3	0.44	0.25	0.04	257	29.70	0.56	7.02	2.12
4-6	0.45	0.23	0.24	360	31.10	0.61	8.77	6.62
18+	0.45	0.26	0.38	379	33.80	0.64	8.89	8.24

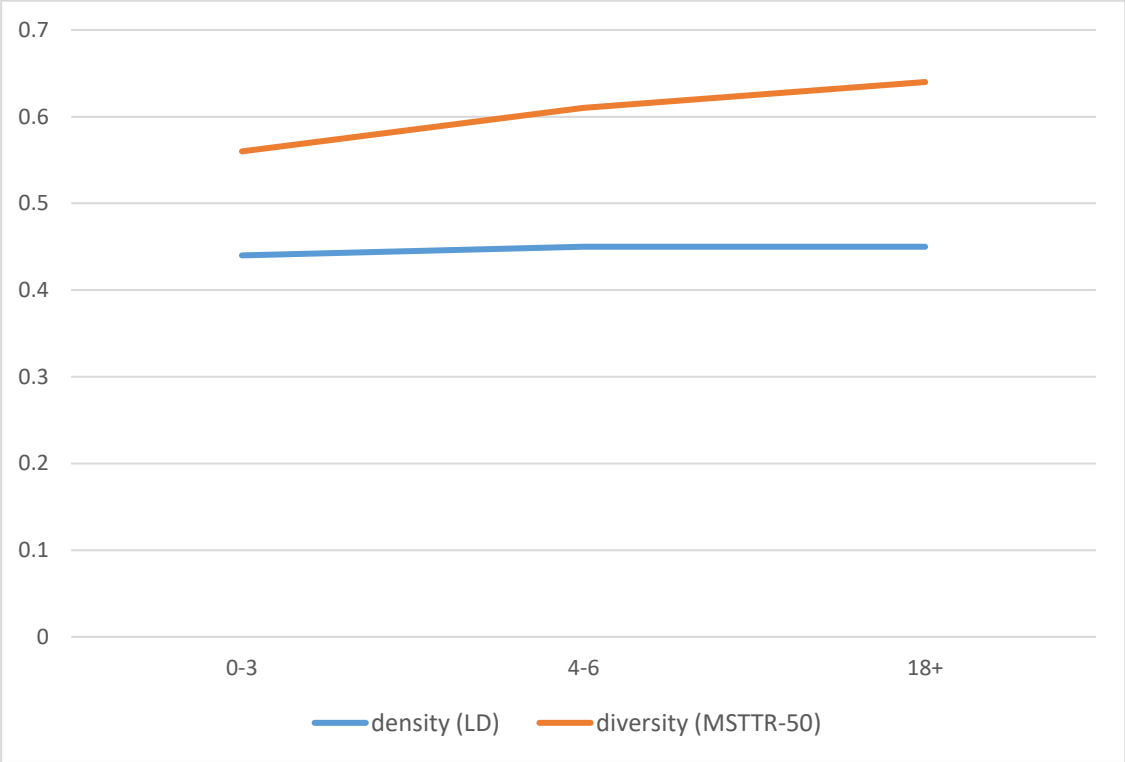
When it comes to lexical sophistication (LS1), the numbers are again similar across age groups. Approximately, there seems to be one relatively rare word (lexical word that does not belong in the top 2000 most frequent words according to the BNC) per 4 lexical words in total across all age groups, that is, one sophisticated lexical word and three common lexical words on average in utterances containing ditransitive constructions. Although it may be redundant to scan for the amount of sophisticated verbs because the utterances centre around a constant ditransitive *give*, the analysis still reveals a spike in sophisticated words with the progression of age (CVS1). Such results imply that with the progression of age, other elements (including sophisticated verbs) are added to ditransitive constructions in the produced utterances (e.g. *give the man a ball or else I'll force you*). Although a measure of diversity rather than sophistication, the SVV1 index concerning verb variation also shows an increase with the progression of age.[41] Not only do 4-6 age groups and adults use less common verbs more often, but they also see to use a wider range of the verbs overall even in sentences revolving around ditransitive constructions. The question is what role does the complexification of a larger language unit such as a sentence bear on the children's processing of the embedded construction, or in this case, what does utterance complexification in the input bears on the children's processing of the embedded ditransitive?

When compared to lexical density, lexical diversity seems to have a somewhat steeper curve as far as utterances containing ditransitive constructions are concerned (see Figure 6). This is not necessarily surprising given that the two indices do not always demonstrate stable correlations amongst themselves and that they measure very different concepts. However, when interpreted together, they tell us where exactly the changes occurred in the samples. Considering that the density remained consistent across age groups and diversity grew, we can

[41] Once again, note that the effects are not brought about by the disproportion in the number of utterances produced by age groups because the values in question were calculated on a random sample of utterances that was equalized across age groups.

assume that diversity was not elevated due to prevalence of either lexical or grammatical words. Instead, the utterances in the speech of children aged 4-6, and then in adult speech, have either grown proportionately in the number of lexical and grammatical words or they have stayed the same in terms of sheer volume, but with an increase in the number of distinct word forms in both of those cases. When we add the calculation of MLUs to this assessment, which tells us that the average length of the utterance did increase with the progression of age, we begin to understand that both length and diversity increased, but the change in lexical and grammatical words-ratio was absent in the case of utterances with embedded ditransitive constructions. An increase similar to the one expressed by MSTTR-50 can also be observed with RTTR.

Figure 6. Density and diversity in the ditransitive give-NP-NP



One of the most reliable indices pertaining to the range of lexical items employed by the speaker has been the number of different words in the sample, especially in the context of child language development (cf. Klee, 1992; Miller, 1991; Moyle et al., 2007; Hadley et al., 2018 etc.). When it comes to utterances centred around ditransitive constructions, both the total number of different words (NDW) and the averaged NDW of 10 random subsamples also show an increase in the variety of lexical items employed with the progression of age. The size of the increase is not particularly large, but along with the measure of sophistication, it indicates where

the increase in variation occurs. Given that the indices of general lexical sophistication were relatively stable (even lexical sophistication which takes into account the types; see the Appendix 10.2), whereas verbal sophistication slightly increased, the data suggests that the age-increased variety (regardless of its size) in utterances containing ditransitives was anchored in some common general lexical items, as well as some less frequent verbs (at least in terms of the entire sentence). In turn, the nature and volume of differences in view of lexical richness suggest that adults (at least as far as the language surrounding ditransitive constructions is concerned) moderate their language use around children. In order to be understood better, the gravity of increase between 0-3 and 4-6 age groups has to be additionally assessed in relation to the other constructions observed in this research.

3.3. In sum

- There is a significant rise in the use of ditransitive constructions with the progression of age. The amount of overall use of ditransitives almost doubles between the 0-3 and the 4-6 age groups, and the increase continues with difference of approximately 30% between the 4-6 age group and adults.
- When it comes to ditransitive constructions, the data shows a significant degree of similarity across age groups in the distribution of both syntactic patterns and lexical items. Although there is a considerable overlap in the proportion of particular ditransitives with respect to their argument structure and complexity, it can also be noted that the overall complexity increases with the progression of age. Children use shorter and simpler constructions, at least in terms of constituents that make up the constructions they produce.
- The overall similarity in the use of ditransitive verbs was noticeable; the verb *give* was found to be the most commonly used, accounting for over a third of all two-object constructions in each age group studied. The verbs *tell, get, show, make*, and others follow *give* in terms of frequency, with only minor differences in their order of usage. The results align with the expected usage patterns based on input and output: (1) the verbs most frequently used by adults are also the most frequently used by children, and (2) although the overall distribution is similar, children's early production of ditransitives tends to concentrate on a smaller set of verbs compared to adult language

(however, it cannot be said that the 0-3 age group concentrates on fewer items in their usage than the 4-6 age group).

- The frequency order of these verbs in ditransitive constructions showed a remarkable degree of similarity across different age groups. The results of the Spearman's rank correlation analysis reveal a strong correlation between the 0-3 age group and the other two age groups, indicating a high and statistically significant similarity in the frequency-based ranking of the most frequent 22 ditransitive verbs used. Interestingly, there appears to be a stronger correlation between the 4-6 age group and the adult group compared to the correlation between the 0-3 age group and the adult group, which contradicts the initial hypothesis.
- When considering the entire range of nouns, the analysis indicates a higher level of similarity between the 0-3 and adult group than between 4-6 and adult group. This disparity is particularly noticeable in the second object position, which is surprising given that the wider range of possible items can occupy the second object slot when compared to the first object.
- The first hypothesis is confirmed across several levels; there is a considerable similarity across age groups with respect to specific nouns, verbs, and syntactic patterns in terms of frequency rankings. However, the hypothesis predicting a greater similarity between the 0-3 age group and adults in comparison to 4-6 and adults is not consistently proven across both verbs and nouns, rendering the data inconclusive in ditransitive constructions.
- The analysis of the data regarding the animacy of the first object in the *give*-NP-NP ditransitive indicates that all age groups predominantly use inherent animates as recipients in this position. The second and third positions, however, show some differences between the age groups in terms of the presence of contextual animates and inanimates. In particular, inanimates are less commonly found in the 0-3 age group, while they are most frequently used in the adult group, with their usage rate doubling as the age groups progress. These proportions suggest that adults are more flexible in their use of inanimates in the first object position compared to children, especially those in the youngest age group. The expectation that the youngest age group would use animate items considered prototypical in designated slots frequently is confirmed as well, but not necessarily to the point where this trend is more pronounced than in 4-6 age group.

Regarding the animacy of the second object in the said ditransitive pattern, the comparison across different age groups surprisingly reveals a more limited utilization of object types based on their animacy. While the recipients predominantly consist of inherent animates, the themes in ditransitive constructions are predominantly composed of (inherent) inanimates.

- In terms of lexical density, there is minimal or negligible variation observed across all age groups in utterances that revolve around the double-object *give* ditransitive construction, and the results are similar for the index of lexical sophistication (LS1). However, when it comes to lexical diversity, there appears to be a slightly steeper progression as observed in utterances containing ditransitive constructions. Though restricted to common general lexical items and some less frequent verbs, the data indicates that the average length of utterances increases with age with respect to both length and diversity. Still the absence of more significant differences could be explained by the fact that adults moderate their language use around children. The hypothesis on lexical richness and complexity is partially confirmed. While it is true that lexical density remains stable across age groups, there is a clear increase in the range of lexical items employed by the age groups.

4. PERIPHRASTIC CAUSATIVE CONSTRUCTION: CORPUS-BASED COMPARISON OF CHILD AND ADULT SPEECH

The subsequent sections in this chapter present the theoretical foundation regarding periphrastic causative constructions (Section 4.1). This encompasses the explanation on how the construction has been operationalized for the purposes of its extraction from the corpora. It also includes a review of prior studies in the field, with a particular focus on research concerning language acquisition and development. The data obtained, the findings from the analysis, and their interpretation are introduced subsequently (Section 4.2).

4.1. Operationalization and previous research

4.1.1. Defining causation

Causation is a complex term in linguistics and causative constructions have often taken on different characterizations by different authors. Again, it is first necessary to recognize the way in which the constructional view of the grammar characterizes the interface between grammar and lexicon, and take into account that grammatical phenomena differ from being completely general to completely idiosyncratic. The term 'construction' is thus generally used to capture the segments of syntax tied to grammar in a conventionalized, but partially idiosyncratic way (Goldberg & Jackendoff 2004). While the traditional lexicalist approaches emphasize the role of words, especially those that behave as lexical heads, constructional approaches insist on the phrasal patterns that in themselves contribute to the meaning.[42] The question then is which patterns do we treat as constructions and which we do not; as stated in Goldberg & Jackendoff (2004), some constructions are fairly easy to spot because of their unusual syntax and/or bits of specified morphology. The main problem in clearly defining the causative construction is its multidimensionality in terms of function and form, i.e. the definitions of causative constructions revolve around its semantic and syntactic features. According to Talmy (2000), causation involves two subevents: the causing and the caused one. According to him, the causing event gives rise to the caused event, meaning that the caused

[42] This does not mean that the word itself cannot represent a construction, i.e. even complex words are theoretically perceived as constructions (see Goldberg, 2006, 5; Booij 2010, 15), which in turn makes a morphological causative a construction of sorts.

event would not happen if there was no causing event and that the caused event did indeed occur.

When referring to the causative construction, like any other, the claim is that the "semantics of the constructional subevent predicts the syntax of the construction" (Goldberg & Jackendoff, 2004, p. 538). Basically, in the case of resultatives, we are distinguishing between the verbal subevent and the constructional subevent. The resultative phrase *Willy watered the plants flat* contains two separable subevents, but the construction is the one entailing the interrelationship between the two. It is not only that *Willy made the plants flat and watered them*, but the inference we draw is that *Willy made the plants flat BY watering them* (Goldberg & Jackendoff, 2004, p. 538). For instance, when the verbs *make* and *get* imply causation – one of their lexical entries is that of "X cause Y to become Z" – the constructional character becomes evident in the fact that the same distribution of arguments can be observed in the construction with that particular causative meaning (such as the sentence *Willy made/got the plants flat*). In their view, this is the proof of the fact that "the semantic argument structure of the constructional subevent determines the syntactic argument structure of the sentence by general principles of argument linking" (ibid., p. 539).

According to Dixon (2000, p. 30), a causative construction entails "the specification of an additional argument, a causer, onto a basic clause", whereby "a causer refers to someone or something (which can be an event or state) that initiates or controls the activity". The causer is specified if a causative construction is a result of a ('valency-increasing') syntactic derivation process (e.g. in the shift from an intransitive to a transitive *John tripped → Mary tripped John*) (ibid., p. 38). Simple causatives "prototypically apply to intransitive verbs and derive transitives", whereby the subject becomes the object in a causative construction (Dixon & Aikhenvald, 2000, pp. 2-3). In this view, the initiator of the activity corresponds to the syntactic–semantic function of a transitive subject (A) and in this example corresponds to *Mary*, while the original S argument (*John*) in the intransitive construction transposes to the O argument in the new transitive construction. The general point is that in the transitive construction the causer always becomes A, while the original A (causee) can be syntactically reassigned (e.g. in the periphrastic *I forced him to go*, causee is the original subject of the verb *go* but is reassigned to the O function of the causative verb in the periphrastic transposition) (ibid., p. 36).

Causation tends to be a primarily semantic mechanism, and only afterwards a syntactic one. This is evident in the effect fundamental to the causative derivation—the introduction of the additional participant (termed as the 'causer') which then syntactically translates to the addition of an argument (Dixon & Aikhenvald, 2000, p. 16). The interpretation of causation as a primarily semantic process goes against the 'syntax first' and 'autonomous syntax' approaches to language. Instead of treating the syntax as an independent process, the emphasis should be on the interrelations between syntax and semantics, i.e. the ways in which one affects the other as well as the chronology of these transformations.

Dixon and Aikhenvald (2000, p. 13) state that the characteristics of prototypical causatives are:

> "(a) Causative applies to an underlying intransitive clause and forms a derived transitive; (b) The argument in underlying S function (the causee) goes into O function in the causative; (c) A new argument (the causer) is introduced, in A function; (d) There is some explicit formal marking of the causative construction."

Furthermore, their claim is that if a certain language has a causative derivation, it will apply to intransitive verbs and form transitives, and the inability of the causer to be included in the underlying intransitive variation of the construction in question.

According to Kulikov (2001, p. 888), a narrower definition of the causatives would entail that they:

> "(i) stand in regular opposition both formally and semantically to the corresponding non-causatives within the verbal system of a given language, (ii) are formally more complex than their non-causative counterparts, and (iii) represent a more or less productive formation."

Depending on the interpretation of what constitutes a prototypical causative construction, linguists consider whether the label is reserved only for those which involve morphological processes, or whether a verb must have a causative meaning. In English, the verb *order* is thus considered more as a speech act verb, while *make* typically entails a causative meaning (Dixon, 2000, pp. 32-33). If one adopts this narrow definition of a causative, sentences like *He was told/ordered to step aside* fall outside of the causative scope. In the thesis, I adopt the same reasoning for the causative construction as for the ditransitive one – the constructions that are taken into account satisfy the terms of a broader definition of a causative construction. From a theoretical standpoint, defining what constitutes a causative construction can be challenging,

especially when the already elaborate semantics of the causative relationship can adopt a number of different forms; for example, consider the distinction between morphological, lexical, and syntactic causatives (see Kulikov et al., 2001). This may prove the task of operationalizing the construction for the purposes of corpus-based exploration more difficult, but the thesis here restricts itself, at least methodologically speaking, to a more straightforward definition anchored in syntax, i.e. the biverbal causative constructions (see Section 4.1.2).

The realization that the causative relationship can be expressed morphologically, lexically, or syntactically has led to the broad usage of the term 'causative construction'. For instance, Goldberg (2013a) frequently makes reference to the causative construction, where it often implies a monoverbal pattern with argument roles of cause, patient and the intermediary (instrument) (ibid., 2013, p. 452). The pattern adopts the label construction precisely because there are argument roles that can be traced and mapped only via construction, i.e. these roles are not verb-specific but construction specific. This is important for Goldberg because such constructional approach reveals both differences and similarities between the caused-motion construction and the non-periphrastic (simple-predicate) causative construction (it becomes clear what can be attributed to syntax and what is there to attribute to the semantics of the verb). On the other hand, the English periphrastic (complex-predicate) causative construction is quite specific on its own, with no competing syntactic patterns that might interfere with delineating research-targeted constructions from the rest (for example, consider the difference between dative and benefactive alternations with double-object constructions; see Broccias & Torre, 2020).

Two defining aspects of the causative construction stem from the characterization of the causative situation itself, and these have been ascertained by Shibatani (1976, pp. 1-2):

> "a. The relation between the two events is such that the speaker believes that the occurrence of one event, the 'caused event,' has been realized at t2, which is after t1, the time of the 'causing event.'
>
> b. The relation between the causing and the caused event is such that the speaker believes that the occurrence of the caused event is wholly dependent on the occurrence of the causing event; the dependency of the two events here must be to the extent that it allows the speaker to entertain a counterfactual inference that the caused event would not have taken place at that particular time if the causing event had not taken place, provided that all else had remained the same."

These two criteria immediately separate the causative sentence from the non-causative one, and consequently a causative construction from the non-causative one. The first criterion is a temporal one; in the eyes of the speaker, the event that is being caused must chronologically follow the event that caused it in the beginning. The second criterion is at the core of the causation concept; in the eyes of the speaker, the second event (caused event) depends on the realization of the first one to the degree that one would not have happened without the other. For the purposes of illustration, Shibatani uses the examples of sentences *I told John to go* and *I know that John went*, which do not constitute causative sentences; the first one does not necessarily lead the speaker to assume that John indeed went after being instructed to, while in the second sentence the speaker's knowing of John's leaving is ultimately inconsequential for his departure. Instead, the causative *I made John go* satisfies all of the aforementioned conditions. Talmy's principle states that the causing event must have something in common with the caused event in order for the causal relationship to hold between two events, and the absence of some common elements negates the possibility of a causal relation between the two (Talmy, 1976, p. 58). Similarly to Talmy, Pustejovsky (1995, p. 186) uses the term "argument coherence" to refer to the same type of relation: "The relation expressed by the causing event and that expressed by the resulting event must make reference to at least one parameter in common", which essentially implies that the two subevents captured by the causative construction must share a participant (for instance, in *I made John go, John* serves as both the patient of the verb *make* and the agent of the verb *go*).

4.1.2. Analytic/Periphrastic Causatives

Kulikov et al. (2001) make a distinction between morphological, lexical and syntactic causatives. The morphological causatives are characterized by an affix which carries the causative meaning. Lexical causatives are represented by verbs which (in themselves) already contain the causative implication – basically, the verb's meaning is "CAUSE Vo", where Vo stands for the embedded base verb (ibid., 2001, p. 887). Finally, there are syntactic causatives which may also be referred to as periphrastic or analytic causatives. It needs to be noted that sometimes there is a confusion with the aforementioned terms, and the label periphrastic, and even syntactic, is used for causatives that are biclausal in nature. In Kulikov et al. they are, in

fact, mentioned as a separate type of causative: "Syntactic causatives are distinguished by many authors from constructions which refer to causative situations but do not represent cohesive units, thus being biclausal sentences" (2001, p. 886). Nevertheless, as they state later on, the distinction between the syntactic causatives and biclausal causatives is not "clear-cut" and it depends on the degree of fusion (for instance, the 'same predicate' causative *make someone go* constitutes a syntactic causative, while the *force someone to go* represents a biclausal causative).[43] Similar typology can be found in Dixon (2000), as he distinguishes between morphological causatives, causatives making up a single predicate, syntactic causatives (periphrastic or analytic causatives), lexical pairs in causative relation and ambitransitive verbs that can be regarded as causatives, and finally the realization of the causative effect through the exchange of auxiliaries which has been observed in some languages spoken by indigenous Australians (Mangarayi and Ngan'gityemerri).

In (English) grammar, the term periphrastic is used for expressions employing two or more words to denote what can usually be replaced by sole inflection (Collins English Dictionary, 2021). In English, the periphrastic causative construction can also be described as consisting of a noun phrase followed by a causative verb and another noun phrase followed by an infinitive, whereby the first noun phrase is the subject of the main causative verb and the second noun phrase the subject of the infinitive (Hollmann, 2003). Causative verbs such as *have* and *make* can take a bare infinitive (*have/make someone do something*), while the verbs *get, cause, force* and others take a full infinitive form (*get/cause/force someone to do something*). Naturally, there are other distinctions to consider within the periphrastic causative, but these are at the lower semantic level. For instance, when it comes to analytic/periphrastic causatives, Stefanowitsch (2001) distinguishes between different types of causal links encoded (i.e. thematic roles of the participants). For instance, *make*-causative can encode both agent-like and stimulus-like participants when it comes to the subject of the main verb (the causer) (ibid., 2001, p. 3):

"Agent: Several hours later she saw police arrive and ***make*** villagers dig the graves.

[43] This research is not interested in the strict separation between the two, and takes both into account when it comes to their acquisition and processing. This distinction between biclausal and monoclausal syntactic causatives is not without its problems, as it seems to rest on the fact that one verb takes a bare infinitive, while the other takes a full infinitive form (for example, consider the *help*-construction which can take both). For a detailed discussion on differences between monoclausal and biclausal structures, and the ways in which these can be tested, see Batinić Angster (in press).

Stimulus: The semi-nude picture of O.J. and Nicole Simpson on the cover of People magazine ***made*** me shudder."

Comrie (1976) argues that the causative sentence is characterized by an underlying structure that consist of a matrix and an embedded sentence and that these are fused together in some languages to form a derived structure, where there is no obvious separation of the clauses. The structure is derived from a matrix sentence that expresses the causer of the action in the position of the subject and the embedded sentence with the subject that takes on the role of the one who carries out the action. Nevertheless, Comrie later on argues that English hardly exhibits structures, where there is a "compression of an underlying complex structure with embedding into a derived-structure simplex sentence" (1976, p. 303). However, biverbal causative constructions in English surely exhibit a considerable degree of fusion as well. For instance, in the derived construction *John made Mary give the book to John*, we have two underlying predicates which can be paraphrased as: *John forcing Mary into something* and *Mary doing that particular something*. As noted by Kulikov et al. (2001), patterns such as *make NP give* can be regarded in many ways as 'same predicate' causative construction. There might be some hesitation from Comrie to use English as an example, partly because sentences such as *I opened the door* qualify for a causative sentence just as *I made John go* (Shibatani, 1976). To a certain extent, one could even argue that the monoverbal *I opened the door* might contain a more complex underlying structure, at least on the conceptual level (thus merging the sentences such as *I did something* and *the door opened*). A subset of verbs which exhibit this kind of causative sense in the transitive clausal pattern are called 'labile verbs', with approximately 800 of them existing in English language (see McMillion, 2006). They can occur in both transitive and intransitive patterns (*I opened the door* vs. *The door opened*), In the transitive clausal pattern, these verbs express causative events where the subject initiates a change in the direct object, acting as an external agent, whereas in intransitive patterns, the subjects, typically inanimate entities, are either interpreted as causing the action themselves or are affected without any external cause being relevant (ibid.). Naturally, it is hard to justify this reasoning in terms of language processing, and this is just a part of the reason why such constructions are excluded from the research.

In order to distinguish between the lexical and periphrastic causatives (for which he uses the label 'analytical', Lemmens (1998, pp. 21-28) uses the example of the verb *kill* which had been previously referred to as a lexical causative. The lexical causatives and the periphrastic

causatives are equated on the basis of semantics – for instance, the expression *cause to die* and *kill* were assumed as semantically equal. The fact that 'kill' derives from the 'cause to die' stems from the generative perspective on semantics – through lexicalization, the verb is thus a derived form of the aforementioned underlying structure. Lemmens claims otherwise even in terms of semantics and claims that lexical and periphrastic causatives "encode different conceptualizations of a given situation" (1998, p. 21). The general definition of the causatives encapsulates verbs that refer to a situation where there is a causal relation between two events, whereby one of the events is believed by a speaker to be caused by another, i.e. causatives refer to verbal constructions that mean *cause to V* or *make V* where the V stands for the embedded verb (Nedjalkov & Sil'nickij, 1973; Kastovsky, 1973; Kulikov, 2001). This general definition encompasses sentences such as *John opened the door*, *Peter made John go* and the morphological causative suffix in Turkish: *Ali Hasan-ı ol-dur-du* (Ali:NOM Hasan-ACC die-CAUS-PAST) meaning *Ali killed Hasan* (Kulikov, 2001; Comrie, 1976).

One of the reasons as to why so much focus as been put on defining causative constructions is in many ways connected to cognition and the way linguists assume how we process language, i.e. the fact that we distinguish between the synthetic causatives and periphrastic constructions is in relation to the history of why a verb *kill* may be, according to some (Lakoff, 1966), treated the same as 'cause to die'. In order to illustrate why the synthetic causative differs from the periphrastic one, Fodor (1970) presents the case of *John caused Mary to die*, analysing the interplay between syntax and semantics via the 'do so' transformation of it. In the transformation examples from Fodor (1970, p. 431), the contention is that the example (8d) is not possible, and yet it should be for the verb 'kill' to be equated with 'cause to die'

(8) (a) *John caused Mary to die and it surprised me that he did so.*

(b) *John caused Mary to die and it surprised me that she did so.*

(c) *John killed Mary and it surprised me that he did so.*

(d) **John killed Mary and it surprised me that she did so.*

The argument about the lack of equivalence between the two can be observed in the comparison between (8b) and (8d). Sentence (8b) is possible in English language, and there is no confusion that the speaker will be aware what the 'do so' construction is referring to. However, this is not the case in sentence (8d), i.e. it is not clear that 'do so' refers to Mary's dying. Another observation by Fodor when it comes to the difference between the monoclausal and the

biclausal causative has to do with the temporal aspect of it. The examples from Fodor (1970, pp. 432-433) illustrate this point:

(9) (a) *Floyd caused the glass to melt on Sunday by heating it on Saturday.*

(b) **Floyd melted the glass on Sunday by heating it on Saturday.*

The point that he makes is that the implication of the periphrastic causative is that one can cause a certain event, or bring about a certain effect, by doing something at a time which is distinct from the particular time of the caused event. If we look at the monoclausal synthetic causative *Floyd melted the glass*, it is not possible to assume this temporal gap between the cause and effect (i.e. if you do melt something, the action of melting happens simultaneously with the melting of the object in question). As Lemmens (1998, p. 25) puts it:

> "At the risk of arguing the obvious, I maintain that lexical causatives such as *kill* almost invariably code events in which the agent's deed and the patient's affectedness fail to coincide referentially. However, unlike analytical causatives (like the cause to-construction), lexical causatives neutralize this difference and code the components as concurrent and integrated."

Many problems that arise from the definitions of constructions relates to the distinction between the so-called deep and surface structure. For instance, this is at the root of the monoclausal vs periphrastic causative discussion.[44] In terms of deep structure, it is safe to assume that the periphrastic causative is a biclausal causative construction, i.e. there is a subject for each of the events entailed by the construction. However, at the surface level, the biclausal construction can be interpreted as a monoclausal one (especially with causative constructions where the causee takes an intransitive verb) containing the complement phrase that may be paraphrased as a clause. For instance, in the sentence *Her friend made Mary cry*, the action chain slightly changes as there is no final patient we usually expect to encounter in the periphrastic causative; *Mary* is thus the patient and the final recipient of the energy initially brought about by the causer. The fact that this is a monoclausal structure is evident in the fact that *Mary* is the object of the (arguably) phrasal verb *make cry*. These concerns about the structure of the periphrastic causative may be resolved by the passivization test (Gilquin, 2010, p. 65), which shows that some verbs such as *have* allow for different interpretations of arguments in terms of the subject-object relations. *Mary was made to cry by her friend* differs from the passivized form of the

[44] While the discussion on definitions is mentioned here, in the context of this research, it is more important to recognize the difference between the surface forms, partly due to the focus on language acquisition, but mostly because of the corpus methodology and the necessity for consistency and clarity in the analysis.

sentence *Her friend had Mary kill Jack*, where even in the passivized *Her friend had Jack killed by Mary*, *Mary* remains the subject of the verb *kill*. According to Gilquin (2010, p. 65), this can be seen as the proof of the biclausal nature of the causative construction with *have*.

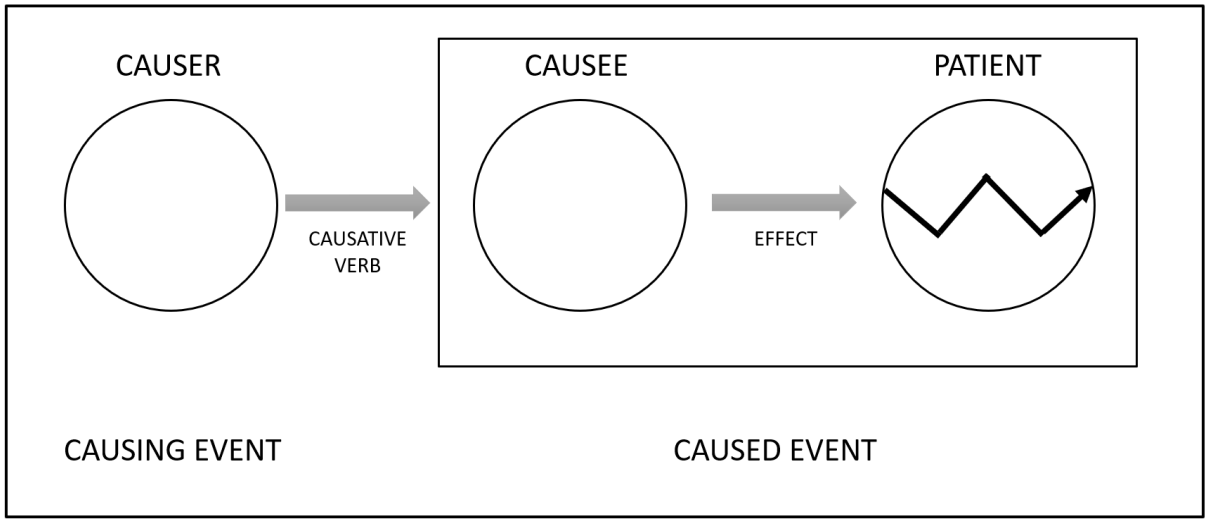
By now, it is important to once again highlight that the thesis focuses exclusively on periphrastic/analytic causatives that take either full infinitives (e.g. *get him to do*) or bare ones (e.g. *have him do*). Another biverbal causative construction is the so-called passivized syntactic causative (whereby two words make up a single predicate with the auxiliaries *have* and *get*; e.g. *get this done*), but this is beyond the scope of this research.[45] Syntactically speaking, the ordering of their constituents corresponds to the form $NP_X \rightarrow V_{CAUSE} \rightarrow NP_Y \rightarrow VP_{EFFECT}$, where the NP_X is the subject of the main verb (i.e. V_{CAUSE}). In the periphrastic variant, NP_Y is the object of the main verb (i.e. V_{CAUSE}) and the subject of the subordinate verb (VP_{EFFECT}), while in the single predicate variant (passivized causative), NP_Y is generally the object of a biverbal predicate. In the periphrastic variant NP_Y thematically corresponds to the agent of the embedded verb, while in the passivized syntactic causatives it corresponds to a patient. In the periphrastic variant NP_Y is the causee, while in the passivized causative construction, the causee is external. In the passivized construction, the causee is removed from the sentence (it can be introduced as the optional by-complement; e.g. *John had him killed by professional assassins*) despite the fact that the causer remains the subject of the sentence. In other words, *John* may be the causer in the sentence *John had him killed* but not necessarily the agent of the embedded verb (causee) (as in: the killing was done by professional assassins) and in the sentence *John got him killed*, again, the source of the killing might be unintentionally related to *John* (as in: John did something that unintentionally led to the killing of one man).

Some linguists (see Dik, 1997, p. 9) have argued that the periphrastic causative construction can be explained in terms of valency extension, meaning that the number of arguments in the source construction can be increased with this additional causativisation; e.g. the transposition of the sentence *The dog ate the food* to *The owner made the dog eat its food*. Gilquin (2010) suggests that within the framework of Cognitive Linguistics, the periphrastic causative construction may be best explained via the addition of the link to the so-called action

[45] The syntax of the passivized causative significantly differs from the periphrastic causative. It exhibits a single predicate as opposed to a complex one in the periphrastic causative construction, with the verbs *get* or *have* taking on the role of the auxiliary verb. Moreover, while there is a distinction in using *have* and *get* as the auxiliary in the passivized causative, the role of the causative verb in a periphrastic causative is pivotal to the meaning of the construction (consider the degree of semantic difference between *force him to do* and *cause him to do* in comparison to *get this done* and *have this done*).

chain. Gilquin (2010) borrows the notion of the action chain from the theory of Cognitive Linguistics, or more precisely, she adapts Langacker's diagram (Figure 7; 1990, p. 221) for the causative construction. The diagram of the periphrastic causative construction graphically illustrates the addition of a link to the beginning of the action chain, which in turn leads to the inclusion of the action source in the predicate of the construction. In such operationalization of the causative construction, the causer becomes the initiator of the action chain, the causee is the first recipient of this energy, and the final recipient is called the patient. In between the causer, cause and the patient, there are two subevents. First is the so-called causing event, formed by the causer and the causative verb. After the energy has been transferred to the causee, another event is initiated – this event is called a caused event in which the causer/action chain initiator needs not be directly involved.

Figure 7. Schematic action chain of a periphrastic causative construction (adapted from Langacker, 1990, p. 221)



Not all verbs with causative implications are necessarily treated as periphrastic causatives, or as verbs constituting causative constructions in general terms. Often, the disagreement about what constitutes the causative stems from the semantics of the verb, and consequently the construction. For instance, Dixon (2000, pp. 22-23) claims that the narrow interpretation of the prototypical causative construction would include a sentence such as *Mary made John go*, but it would not include the sentence *Mary ordered John to go.* The wider interpretation, which is adopted in this research, permits verbs such as *order, persuade, convince* and others to be treated as non-prototypical causatives. However, despite the fact that

some constructions doubtlessly convey a causative relation, the line has to be drawn somewhere. For instance, Song (1996, p. 36) regards the sentence such as *I speak and child eats* as a causative construction, but although we can semantically map out a causative relation similar to the one of ordering someone to do something, the matter is approached from the perspective of language processing and acquisition which warrants a coherent and precise definition of the studied phenomena, The research for this thesis is thus restricted to a selection of verbs, partly due to methodological concerns and the difficulties related to the extraction of the targeted constructions from the corpus.

Rather than assuming a monoverbal, biverbal or even a biclausal description of the periphrastic causative construction, from the general perspective of our research, it may be best to define it as a "two-part configuration" where a non-finite complement clause is controlled by the causative verb, where the relationship between the two implies a causal relation that brings about a certain effect in the end (Gilquin, 2010, p. 1).

Wolff & Song (2003) argue that beside the straightforward implication of the verbs that are consistently treated as causatives (e.g. *cause, make, force*) precisely because they encode causation in its purest form, the periphrastic causatives encompass a wider category of semantic restrictions. Here, the authors add the concepts of 'enable' and 'prevent'; for instance, the verbs such as *enable, let* and *allow* would also constitute a periphrastic causative that resonates with the semantics of 'enable', while the verbs *prevent, block* and *restrain* belong to the semantic category of 'prevent'. This definition of the periphrastic causative is the one where the "embedded clause encodes a particular result", i.e. there is a strong implication that the encoded result will occur, or a result in a sense of intervention to prevent a particular change that would otherwise occur had there not been an intervention (as in the case of 'prevent') (Wolff & Song, 2003, pp. 285-286). Some authors studied the periphrastic causative construction, limiting themselves only to verbs such as *make, have, get, cause* and *let* (Ammon, 1980; Baron, 1977; Shibatani, 1976). The usual verbs that are mentioned in the periphrastic causative construction have been expanded by Wolff & Song (2003), who came up with 49 periphrastic causative verbs in their analysis of causation based on both syntactic and semantic criteria. They divide the verbs into two groups depending on whether they can be directed towards sentient and non-sentient causees (1), or exclusively sentient causees (2):

(1) *allow, block, cause, enable, force, get, help, hinder, hold, impede, keep, leave, let, make, permit, prevent, protect, restrain, save, set, start, stimulate,* and *stop*

(2) *aid, bar, bribe, compel, constrain, convince, deter, discourage, dissuade, drive, have, hamper, impel, incite, induce, influence, inspire, lead, move, persuade, prompt, push, restrict, rouse, send,* and *spur*

When it comes to the causative to-infinitive verbs in English, Sugawara (2016) has observed their differences and subdivided them according to semantics. Depending on the semantic perspectives of implicativity and affectedness of objects, this lexically-oriented division yields four types of the causative to-infinitive pattern. The constructions with implicative verbs imply that the action denoted by the infinitival complement has actually been carried out, whereas the affectedness of the object reveals whether the influenced participant undergoes a psychological change. If the verb is an implicative one, it means that the action has indeed been carried out. Finally, the distinction is between the implicative verbs which induce a psychological change (1), the implicative verbs which do not necessarily induce a psychological change (2), the non-implicative verbs which induce a psychological change (3) and the non-implicative verbs which do not necessarily induce a psychological change (4):

(1) induce-type: *induce, influence*, etc.

(2) force-type: force, *compel, oblige, drive, cause*, etc.

(3) persuade-type: *persuade, convince, decide, determine*, etc.

(4) order-type: *order, command, instruct, direct*, etc. (Sugawara, 2016, p. 15)

To sum up, the present study delves into periphrastic causative constructions. These constructions are specifically defined as those where causation is expressed through complex predicates that embed either full infinitive or bare infinitive forms of the caused event verb. To elaborate further, syntactically, these constructions follow a specific order of constituents ($NP_X \rightarrow V_{CAUSE} \rightarrow NP_Y \rightarrow VP_{EFFECT}$): the subject of the main verb (referred to as NP_X) followed by the causative verb (V_{CAUSE}), then the subject of the embedded verb (NP_Y), and finally the embedded verb itself (VP_{EFFECT}). It is worth noting that this research adopts a broader definition of periphrastic causatives, which not only includes the prototypical cases but also encompasses borderline instances where the causer simply facilitates the action of the embedded verb rather than directly causing it in a traditional sense (such as *help* and *let*-causative constructions among others).

4.1.3. Previous Research

4.1.3.1. Causative alternation and periphrastic causatives

Much of the research in the area of causative verbs has been focused on the interpretation of causative alternation (Levin & Rappaport Hovav, 1994; Pye & Levy, 1994; Marcotte, 2005; Schäfer, 2009; among others). The phenomenon has also been under the scrutiny of those working in the field of early language acquisition and psycholinguistics (Pinker 1989). The alternation can be illustrated by the transitive-intransitive sentence pair: *Pat opened the door / The door opened*. Some find the alteration interesting in terms of crosslinguistic existence of the phenomenon and its implications. Other are more focused on the theoretical issues around discerning which verbs are inherently dyadic and which are monadic. More precisely, the question is often "whether all intransitive verbs have transitive counterparts with the paraphrase appropriate to the causative alternation" (Levin & Rappaport Hovav, 1994, p. 42), and the results seem to suggest that the intransitivity of a verb is not necessarily sufficient to ensure the verb's participation in the alternation. Some suggest that agentivity is one of the crucial factors in the causative alternation and that the agentive rules might not license the alternation despite the existence of a number of exceptions to the rule (ibid., p. 39). Instead of agentivity, the related notions of 'internally' and 'externally caused' eventualities may help with the understanding of the causative alternation mechanism, and according to Levin & Rappaport Hovav (1994, p. 49), the concept of internal cause subsumes agency. As they further explain:

> "For agentive verbs such as *play, speak, or work*, the inherent property responsible for the eventuality is the will or volition of the agent who performs the activity. However, an internally caused eventuality need not be agentive. For example, the verbs *blush* and *tremble* are not agentive, but they, nevertheless, can be considered to denote internally caused eventualities, because these eventualities arise from internal properties of the arguments, typically an emotional reaction." (1994, p. 49).

Moreover, when it comes to the lack of the causative alternation for an internally caused verb, the answer resides in the properties of the argument structure, meaning that – in order to introduce an external cause – one has to "express the causative use of internally caused verbs periphrastically" (instead of saying *The clown laughed me*, the expression will be *The clown made me laugh*; more examples can be found in Levin & Rappaport Hovav, 1994, p. 55).

Particularly detailed work on periphrastic causatives came into prominence in the last two decades with researchers such as Stefanowitsch (2001), Hollmann (2003, 2006), Gilquin (2004, 2010) and Lauer (2010).

Gilquin's study (2004) shares a part of the methodological framework with this thesis. In her study of the main English causative verbs, Gilquin (2004) also dealt with the periphrastic causative constructions with the verbs *get, have, make* and *cause* and based her study on the corpus data (more precisely, a subset of the British National Corpus), thus orienting herself towards a usage-based cognitive approach to the interpretation of the structures and choices of particular causative verbs. Gilquin (2006) also argued that periphrastic causative constructions accommodate fewer number of secondary verbs than typically assumed, that is, verbs heading the periphrastic causative constructions are apparently restricted in number, but those occurring alongside as collexemes are not as obviously confined. The data she extracted from the British National Corpus revealed several things: the periphrastic causative constructions with *make* as its head verb and bare infinitive form for its secondary verb ("X make Y V_{inf}") were clearly the most frequent ones when compared to alternatives with *get, have and cause* as their head verbs. On the other hand, when it comes to coreferential causative constructions,[46] *get* was the most frequent causative verb heading the periphrastic causative (see Gilquin, 2007). Furthermore, Gilquin (2006) found that, for each of the constructions with the aforementioned head verbs – there are different collexemes that tend to be the preferred choice by the speakers; for example, in "X make Y V_{inf}", the most distinctive verb-collexemes were *feel, laugh, look, think, wonder, appear, seem, want, sound, jump* etc. Gilquin further concluded that the variation in preferences towards certain verbs in the secondary non-finite position suggest that causative constructions should be treated as separate entities, rather than "(near) synonyms" (ibid., p. 40), with respect to their syntax and causative verbs heading the construction.

Similarly to Gilquin, Hollmann (2006) also used The British National Corpus in order to obtain the results for the active and passive periphrastic variations of the causative *make* (also dubbed the most general causative in terms of semantics). Being interested more in the processes behind passivization, Hollmann's aim was to gather the quantitative evidence demonstrating correlations between the properties of transitivity and passivisability. Due to his background in the typological work, he was interested in the correlations that can be stated as

[46] Coreferential causative constructions are those where the causer and the causee constitute a single entity, or two parts of a single entity (e.g. *He got himself to start training again*).

implicational universals. The questions addressed in the work primarily revolve around dissimilarities in passivisability of different English periphrastic causatives, such as *cause, make, get, have, force, persuade*. Passivisability was also explored by Gilquin, who interestingly found that the verb *make* passivises in more than 8% of cases (Gilquin, 2004, p. 256), while the percentage is significantly lower for the verb *cause* and revolves around 5% (Gilquin, 2004, p. 257). In Stefanowitsch's view (2001), passivisability is a feature anchored in the semantic compatibility between the passive construction and the causative construction, whereby the salience of the causer is central for the licensing of the passivization process. For instance, the verb *have* takes a relatively salient causer and because of it appears less suitable for passivization (e.g. *He was had to jump over the fence** vs. *He was made to jump over the fence*).

Hollman (2003) was interested in the rise of the periphrastic causatives *have*, *get* and *make*. In his review of the phenomena, he takes into account psychological factors, synchronic and diachronic approaches to these constructions, and also makes remarks on the usage-based accounts of the constructions (more precisely, the causative *get*). As others, he also considers the syntax-semantics interface of the constructions, and relying on the tradition in cognitive linguistics, asserts that constructions are not merely syntactic templates, but create form-meaning pairings. The conventionality of the meaning in these periphrastic constructions is also considered as a step-away from previous models of language acquisition such as generative grammar. Lauer (2010) was also focused more on the semantic aspects of the construction. Searching for both commonalities and differences in meaning between periphrastic *cause* and *make* (but also touching upon other semi-causatives such as *let* and *allow*), he addressed the dichotomy between the semantic categories of necessity and sufficiency whilst shedding some light on notions such as 'direct causation'.

4.1.3.2. The acquisition of periphrastic causatives

Important data concerning the acquisition of periphrastic causatives was put forward by Snyder and Stromswold (1997), who looked into acquisitional frames for 'causative/perceptual constructions',[47] along with other dative constructions in English (such as 'double-object' and

[47] The category entailed a group of periphrastic constructions with the verbs *hear, make, see* and *watch*, all of which take postverbal nouns followed by predicate constructions (as in periphrastic causative constructions considered in this thesis), and all of which were expected to have a high frequency in the English language.

'*to*-datives'). The authors conducted a research study on the age of acquisition of different sentence constructions in children. The study found that children aged between 1;8 and 3;10 acquired causative/perceptual constructions at a mean age of 2;4.9. The gap in time between the acquisition of *to*-datives and causative/perceptual constructions varied from 0 to 8 months. Among the children they studied, 8 children learned double object datives before causative/perceptual constructions, 2 children learned causative/perceptual constructions before double object datives, and 2 children learned both at the same age. The study also found that children acquired causative/perceptual constructions 2.0 months earlier on average than *to*-datives, with 8 children learning causative/perceptual constructions before *to*-datives, 3 children learning *to*-datives before causative/perceptual constructions, and 1 child learning both at the same age. Snyder and Stromswold's (1997) study indeed found a strong correlation between the age of acquisition of causative/perceptual constructions and the age of acquisition of both *to*-datives and double object datives, but there were no significant differences in the order of acquisition between causative/perceptual constructions and either type of dative (double-object and *to*-datives). Although they could have argued that causative/perceptual constructions are acquired concurrently with *to*-datives, rather than double-object datives, the lack of clear ordering between causative/perceptual constructions and *to*-datives was likely due to the lower frequency of causative/perceptual constructions in adult English, relative to double-object datives. Finally, the authors (ibid.) concluded that the study provided strong evidence for a parametric[48] model of children's acquisition of datives and related constructions.

Additional data on the acquisition of causative constructions comes from two studies on elicited speech (Brooks & Tomasello, 1999; Brooks & Zizak, 2002; cf. Ambridge et al., 2013). The two studies cantered themselves around the prediction that if a verb appears more frequently in a specific construction in the language input, such as the periphrastic causative with the verb *disappear* in *X made Y disappear*, then children will be less likely to make errors

[48] According to Snyder (2021, p. 875), the parametric approach to syntax acquisition in children implies that syntax is highly constrained. The model includes principles, which are syntactic characteristics that we expect to be true in all natural languages, and parameters, which are specific and narrowly restricted points of variation. This approach does not rule out the possibility of some aspects of syntax being idiomatic or restricted to certain words. However, it entails that a significant portion of a native speaker's syntactic knowledge can be accurately characterized using general principles and parameters that apply to the language as a whole. Interconnectedness, which refers to the idea that certain constructions and mechanisms are linked by an abstract parameter, is a key characteristic that we expect to observe during acquisition. For instance, if a parameter connects certain constructions, it makes sense that children learning the language will produce these structural patterns as a cluster in close temporal proximity.

when using the verb in a related construction where it is not allowed, such as transitive causative ones (e.g. *X disappeared Y*). Brooks and Tomasello (1999) suggested that if a child hears the sentence *The magician made the rabbit disappear*, instead of *The magician disappeared the rabbit* (which would be expected since direct causation is involved), the child may infer that the verb *disappear* does not occur in a simple transitive construction. This is because the adult seems to be purposely avoiding using *disappear* in this way and instead using the more marked periphrastic causative construction. Essentially, the repeated exposure to *disappear* in periphrastic causative constructions that involve direct causation prevents the child from using *disappear* in a simple transitive construction. If this type of situation happens frequently, the child may tentatively assume that the verb cannot be used in that construction. In many cases, entrenchment[49] and preemption[50] can work together, as a verb that is strongly established in one usage is not used in another way, and an alternative is used instead. Both Brooks and Tomasello's (1999) and Brooks and Zizak's (2002) study provided support for this prediction in children aged 4-5 years old and older, while the results for younger children were far less indicative of their initial assumptions. In these studies, children were taught a new verb in intransitive sentences, where the experimental group heard the verb used in periphrastic causative sentences and the control group did not. As expected, the group which received the input with periphrastic causative use was less likely to generalise the verb into the transitive-causative construction than a control group, indicating that frequent exposure to the verb in the periphrastic causative made children more cautious about using it in other constructions.

While their predictions were confirmed for older children, there was an obvious mismatch between younger children (two and four year olds) and older children. For example, Brooks and Tomasello (1999) found that 62.5% of the older children used the periphrastic causative construction, compared to only 37.5% of the two-year-olds. The two-year-olds used

[49] Individuals may be more likely to resist overgeneralising higher frequency verbs than lower frequency verbs with similar meaning and argument structure constraints. This phenomenon has been explained through entrenchment, which suggests that people prefer to use verbs in ways they have heard before, with the dislike for novel uses increasing with overall verb frequency (see Robenalt & Goldberg, 2015). Much like entrenchment, preemption is based on a similar idea whereby the frequency in the input is linked to one's flexibility with the item's usage.

[50] To differentiate between preemption and entrenchment, consider the example of the error *The magician disappeared the rabbit*. Preemption only occurs when this verb is used in its closest synonym construction, such as the periphrastic causative (e.g. *The magician made the rabbit disappear*). On the other hand, entrenchment occurs when this verb appears in any construction or even as a standalone utterance (e.g. *He disappeared.; Would you like me to disappear?; Disappear!*). Therefore, preemption is determined by the frequency of a verb in a single construction, while entrenchment is determined by the overall frequency of a verb (Ambridge et al., 2013).

simple transitive and intransitive constructions more frequently than the passive and periphrastic causative constructions; the authors attributed this to the fact that the latter constructions were considered structurally more complex and less frequently used in the speech children hear. In contrast, the older children used the constructions observed more evenly. They explained that the reason younger children (aged two and four) did not demonstrate improvement in their ability to use different construction types correctly was because they did not connect the information about transitivity across different construction types. Essentially, the younger children may not have realized that a verb used in a periphrastic causative construction has the same transitivity as a verb used in a simple intransitive construction, and similarly, a verb used in a passive construction has the same transitivity as a verb used in a simple transitive construction.

The study by Brooks and Zizak (2002) built upon the study by Brooks and Tomasello (1999) by upgrading the experiment and modifying the experimental and control groups in order to avoid confoundment between preemption and entrenchment. Unlike the previous study, all groups were equally presented to novel verbs in simple transitive and intransitive constructions. Brooks and Zizak (2002) found that six to seven-year-old children can benefit from hearing a verb used in alternative constructions that do not change its transitivity in order to infer its transitivity status. However, their study also did not show a similar benefit for four-year-olds exposed to novel verbs in preempting alternative constructions, suggesting that it may take several years for children to develop an understanding of how different argument structure constructions are related in terms of their transitivity. The researchers do not claim that children never relate information about usage patterns across verbs, and suggest that some kind of structure mapping or similarity-driven abstraction may play a role in forming abstract linguistic constructions. However, there is still much to learn about how input shapes children's use of argument structure constructions and how information processing constraints affect distributional learning in children. The study demonstrated that children selectively use indirect negative evidence to constrain their use of novel verbs, and the distributional statistics used by the participants are considered crucial for language comprehension and production in adults.

In the study by Ambridge et al. (2014), both 3-year-olds and adults were given sentences with ungrammatical noun-verb-noun structures containing intransitive only verbs (e.g. Bob laughed Wendy) and were asked to choose between construction-meaning and sentence-repair interpretations. The majority of participants from both age groups (at least 82%) chose the

construction-meaning interpretation (e.g. *Bob made Wendy laugh*) rather than the non-causal sentence-repair interpretation (e.g. *Bob laughed at Wendy*), which, according to them, supports the idea that construction-meaning plays an independent role in adult grammar and child language acquisition. While the research did not suggest anything on whether these mappings were learned entirely from input or if they had some innate basis, the findings indicated that participants tend to use construction semantics to interpret ungrammatical sentences even if it requires overriding their knowledge of verb meanings and argument-structure restrictions. According to Ambridge et al. (2014), any successful model of adult grammar and child language acquisition must take into account the role of construction semantics. In the subsequent study, Ambridge et al. (2015) conducted a study on 216 participants separated into three age groups (5-6 years, 9-10 years, and 18-22 years) who were asked to rate the acceptability of sentences with verb argument structure overgeneralisation errors on a five-point scale. The aim of the study was to test two proposed solutions to the question of how children avoid such errors. The first hypothesis, preemption, predicted that the acceptability of the error would decrease with the frequency of the verb in its most similar construction. The second hypothesis, entrenchment, predicted that the acceptability of the error would decrease with the overall frequency of the verb. The results showed support for the entrenchment hypothesis for the older age groups, but not for preemption. This suggests that retreating from errors is not achieved through specialized mechanisms, but rather through a process of construction competition that is probabilistic. In conclusion, entrenchment appears to be a robust solution to the retreat from error across different error types.

The study by Robenalt and Goldberg (2015) explored how speakers determined when they could use language creatively and when they could not. The paper investigated whether verb frequency was always relevant in determining the acceptability of novel sentences, or if it only mattered when there was an available alternative way to express the intended message with the chosen verb, which was predicted by statistical preemption. Two experiments were conducted where participants rated novel uses of high and low frequency verbs in argument structure constructions in which those verbs did not usually appear. The results revealed that the dispreference for a novel use with a high frequency verb relative to its lower frequency counterpart was stronger, but only for sentences with an agreed-upon preferred alternative phrasing that would compete with the novel use for acceptability. When there was no consensus about a preferred way to phrase a sentence, verb frequency did not have a predictive effect on

sentence ratings. The authors interpreted this to suggest that while people generally prefer familiar formulations to novel ones, they are willing to extend verbs creatively if there is no readily available alternative way to express the intended meaning.

As far as more recent research is concerned, Ambridge et al. (2020) conducted a study that examined how children learn language rules while avoiding errors. They tested three different proposals using participants from three age groups (5-6 years, 9-10 years, and adults) and five languages (English, Japanese, Hindi, Hebrew, and K'iche'). The participants rated the acceptability of grammatically correct and incorrect sentences that described causation events. Apart from verb-semantics hypothesis, they tested entrenchment hypothesis, and preemption hypothesis. The entrenchment and preemption hypotheses predicted that differences in acceptability ratings would be based on the verb's relative overall frequency and frequency in nearly-synonymous constructions. The study found that entrenchment/preemption effects were observed for all age groups and languages except K'iche'. All languages showed effects of event-merge semantics, except K'iche', which showed only effects of supplementary semantic predictors.

Bidgood et al. (2021) used different methods to investigate how children and adults avoid argument structure overgeneralisation errors, such as *You giggled me*. They examined how semantic and statistical constraints, including preemption and entrenchment, influence judgments of grammatical acceptability of 120 verbs in transitive and intransitive sentences for all ages. They also looked at syntactic priming as a generator of overgeneralisation errors in children aged 5-6, to investigate whether the same constraints operate in production. The results showed that preemption, entrenchment, and semantic effects are real and affect judgments for all ages, but only preemption was observed in production for transitivisation errors with intransitive-only verbs. The researchers concluded that while the effects of preemption, entrenchment, and semantics were real, some aspects of the experiment may have obscured their influence.

To sum up, the majority of research on (periphrastic) causatives in the context of language development and acquisition has looked into errors children produce in terms of both syntax and semantics, and the potential reasons for these errors. The explanations put forward revolve around competition as the main predictor of children's production and understanding. A good synthesis of the cluster of these findings that can be interpreted in view of language acquisition

has been provided by Ambridge and Lieven (2015). As they succinctly put it, children make errors in language acquisition because of the competition between different forms of language, from single words to abstract constructions; in the same way, the mechanism of competition that is responsible for errors is also responsible for why children stop making errors (ibid., pp. 494-495). Early on, children may use the wrong form because it is more frequent or because they have not yet learned an appropriate alternative. However, as they learn new schemas and constructions that better fit the items in their intended message, they retreat from error. This process is influenced by two factors: statistics and semantics. Children learn the probabilistic links between items and constructions, causing certain constructions to be activated more frequently. Additionally, they learn the fine-grained semantic properties of different construction slots, making certain verbs more appropriate for certain constructions over others. As a result, the likelihood of children accepting and comprehending verbs in non-attested constructions decreases as verb frequency increases and semantic compatibility with the target construction improves.

4.2. Periphrastic causative constructions – results

4.2.1. Extraction method and the examples of periphrastic causatives

Given that I am working on the already existent corpora, the methodology of the extraction process relies on the POS (Part-Of-Speech) tagging provided by the platform Sketch Engine. Essentially, the methodological process comes down to clearly defining the targeted construction and employing the syntax that will locate the constructions of interest in the corpora. Drawing from the definitions of the periphrastic causative construction stated above, I looked at the standard periphrastic causatives with the verbs *get, make* and *have,* but I have also expanded the selection of these verbs with the ones that take full infinitives, and carry a compelling causative meaning such as *cause, force, compel, coerce, lead, oblige, drag, drive, press, pressure, blackmail, strong-arm*. The category is further expanded with the verbs *order, command, appoint, direct, require, request, demand, dictate, ordain, convince, motivate, urge, ask, beg, bid, tell, instruct, authorize, empower, entitle, endorse, sanction, support, insist, talk*; these are verbs which do not necessarily imply the realization of the outcome (the final outcome of the caused event), but there is a strong implication of the realization (whereby the causee

does what the causer led them to do). Further considered are the verbs in the category of allowing someone to do something – such as *allow, permit, enable, grant, license, approve*. Furthermore, the productive *let* and *help* were also considered separately due to their productivity and frequency. They do not include the prototypical semantic information conveyed by the causative, but they do share a considerable portion of syntax and semantics with causative verbs.

Examples of negative causation were also considered, but ultimately were not included in the extraction process and these examples might prove revealing in some further research. These constructions can be described as 'V_x someone/thing from V-ing something' (where the V_x can be replaced by verbs such as *forbid, prevent, stop, ban, block, deny, cancel, censor, disallow, halt, hinder, impede, inhibit, oppose, restrain, restrict, avert, forestall, prohibit, thwart, limit, suspend, disrupt, stall, obstruct*).

All in all, the analysis aimed to cover the majority of periphrastic causative constructions, even those which are semantically not clear-cut in terms of causation, i.e. they all imply it to a higher or lesser degree, or at least imply that the action of the subject (causer) is somehow related to the realization of the secondary event/action (for example, *help* and *let* constructions). These constructions were included primarily because they share the syntax of the rest of canonical periphrastic causatives (just as benefactives were included in Section 3, given that they share a non-prepositional pattern with canonical ditransitives).

Some of the periphrastic causative constructions extracted after filtering out the corpora can be found in Table 15. The examples include a variety of realizations, differing in terms of causative verbs, animacy of the causee, infinitival alternations (*to* vs. bare infinitive) etc. The simplest constructions are restricted solely to the causative verb, the adjacent causee and the effect expressed by the secondary verb (such as *Let me kick*). Others are more complex and, aside from the aforementioned, include the causer and the patient as well (e.g. *Somebody's helping him hold it because the mine's real deep*). As already explained, various verbs have been included as potential verbs expressing causative relationship, and as a result cover instances with *let, ask, help* and others[51] (e.g. *Want to [ask him to find other one of these]* or *[Somebody's helping him hold it] because the mine's real deep*). Furthermore, we find various sentence length in relation to the periphrastic causatives. Some examples, especially as the children grow older, include elaborate sentences, where the periphrastic causative constructions

[51] See the subsequent analysis for more details.

constitute a minor sub-clausal piece (e.g. *I would like to sleep over her at her house every day because she [lets me stay up late] about ten o'clock or twelve thirty*). Some sentences even contain multiple periphrastic causatives (e.g. *And [I command Mark to come in] and hyp [I command Mark to be hypnotized by me and come in and be sliced]*), although this may be due to inadequate inter-clausal separation, where the two clauses actually represent distinct utterances rather than a complex sentence.

Table 15. Examples of (candidate) periphrastic causative constructions found in the CHILDES corpora

Corpus	Age (y; m; d)	Produced causatives
MALAKOFF	1;11;17	*Let me kick.*
WELLS	1;11;30	*I let you mend it.*
HIGGINSON	2;10	*Be right back [I'm get my woolie blanket to keep me warm].**
MANCHESTER	2;10;29	*Can't get the dolly to work.*
DEMETRAS-WORKING	2;2;24	*Have this get out.**
VALIAN	2;5	*Want to [ask him to find other one of these].*
DEMETRAS-WORKING	2;7;4	*[Let me play] and you can sing the grand old flag.*
WEIST	2;9;2	*I gotta go get a chair to talk to you.**
FLETCHER-THREE YEAR OLDS	3;0;24	*He's got a beaker to drink out.***
GARVEY	3;1	*Let me drive the car.*
WEIST	3;4;18	*[Let I drive it here] with my new car with Emma.*
WEIST	4;3;14	*Because [they have water and duckbills swim].***
KUCZAJ	4;3;15	*[Somebody's helping him hold it] because the mine's real deep.*
GLEASON	4;5	*You told [you told you to play that] at dinner time Daddy.*
WARREN	5;10	*[…] and it got um it made the water go up to the bricks and [made it overflow].*
GLEASON	5;2;10	*If I need help you help me.***
CARTER	6	*I would like to sleep over her at her house every day because [she lets me stay up late] about ten o'clock or twelve thirty.*
MACWHINNEY	6;0;27	*And [I command Mark to come in] and hyp [I command Mark to be hypnotized by me and come in and be sliced].*

*dubious but retained in the analysis **removed from the final dataset

Table 15 also includes examples which have been flagged as periphrastic causatives initially, but later removed after the second revision of data. For example, the sentence *He's got a beaker to drink out* was initially kept, but after further scrutiny, it becomes clear that the two verb phrases are not related in the same way as in periphrastic causative construction (forming

a complex predicate), i.e. the noun 'beaker' is not actually a causee, but rather a theme which undergoes the action (the non-finite clause underlying the purpose can be paraphrased as *in order to drink out*).[52] The sentence *Because they have water and duckbills swim* was initially overlooked and included as a periphrastic causative because the *water and duckbills* was wrongly identified as a single entity forced to *swim*, but is actually a complex sentence with two independent clauses and predicates. In a similar fashion, the example *If I need help you help me* was removed from the final dataset; the sentence was initially registered as a periphrastic realization *help you help me*, although in reality represents a conditional sentence where it is necessary to put comma in between *help* and *you*.

There are several examples where the causee is an inanimate noun (*I'm get my woolie blanket to keep me warm* or *I gotta go get a chair to talk to you*), where the thematic roles of the instrument and the causee often overlap, but these cases remain the minority of examples retained in the final dataset (see Section 4.2.2 for further analysis). In the first sentence, the 'woolie blanket' is arguably the instrument but it also performs an action, whereas in the latter case, the 'chair' can either be interpreted as one that is capable of speech, or simply an instrument which facilitates the action performed by the speaker. There are also examples which have been problematic to interpret, but were still retained due to shared syntax with the periphrastic causatives. Aside from *I gotta go get a chair to talk to you*, the example *Have this get out* was also retained in the final dataset, despite the fact that the construction has a twofold possibility interpretation-wise. In other words, it is possible that the sentence is in fact, one with two independent clauses which should have been separated by a full-stop, or that the speaker truly intended to produce a form of periphrastic causative, whereby (in a more loose interpretation) someone was asked to take out something.

[52] The *beaker* can also be interpreted as the instrument which facilitates the process of drinking, thus resembling the conventional thematic relations captured by the periphrastic causative. Still, the example is different from other periphrastic causatives where the inanimate causee was ultimately kept. In the sentence *Be right back I'm get my woolie blanket to keep me warm*, there is a cohesion between the thematic roles of an instrument and the causee, but the example is retained in the final dataset because the sentence clearly denotes the patient (*me*), as opposed to the *He's got a beaker to drink out*, where the two may overlap.

4.2.2. Structural complexity and similarity

When the overall distribution of periphrastic causatives is observed across age groups, the *let* construction is by far the most frequent one (see Table 16; for the corresponding examples, see Table 17). Part of the reason is that it includes the phrasal use of the verb, which semantically may barely qualify for the periphrastic causative, but pragmatically does not really imply causation in any degree (for instance, *let me explain* or *let me introduce myself*). This also relates to the fact that the most common realization in all age groups – *let*-ProPers-V – is considerably more frequent than *let*-N/MOD-N-V.[53] The second most frequent pattern across all age groups is *make*-ProPers-V (e.g. *make him do it*), followed by the third – *help*-ProPers-V (e.g. *help him do it*). Other relevant patterns, in terms of sheer frequency, include various causee-based alternations with the aforementioned verbs, as well as less common patterns with *get*. The least common patterns are those with demonstrative pronouns in causee positions, as well as lexically more specific infinitival patterns with verbs that need not necessarily belong to traditional interpretations of periphrastic causatives. For instance, when it comes to *TALK*-NP-*into*-V, only six instances were observed in the adult group (making it insignificant when normalised per million tokens as can be seen in Table 16 (freq.<0.5).

Among the more interesting observations is the fact that periphrastic causative constructions are equally represented in the 4-6 age group and adult child-directed speech. This is not a tendency that was observed on ditransitive constructions, thus inviting questions as to why this might be the case. While the cross-examination of usage-frequencies regarding constructions is addressed in Section 6.1, it is worth noting that the overall frequencies of this particular construction use are basically the same, the chances for which must also be statistically low. So, the question is, what exactly can we infer from the sheer total frequency of the causative constructions? Does this mean that the language of children in 4-6 age group comes closely to that of adults regarding the overall use of relatively complex syntactic patterns? Does it mean that adults' speech, when simplified for the purposes of conversing with children, starts to resemble the speech of 4-6 age groups? Naturally, there are some discrepancies in terms of individual patterns and verbs heading the periphrastic causatives

[53] Note that the bare 'V' label does not necessarily imply that the construction, or the sentence, ends with the verb. Instead, the V label is used to highlight the fact that the pattern excludes instances which take the full infinitive form (which could be misinterpreted had the VP label been used). Full infinitives are designated as '*to*-V'.

which might shed some light on the aforementioned questions, but there is no doubt that the causative construction production-related obstacles have been overcome by children 4-6 age group to a surprising degree. There is obviously a lot of similarity between 4-6 age group and adults, even when the individual patterns are taken into account, i.e. the most pronounced differences seem to be in some variations with the verbs *get*, *let* and *make*. In terms of percentages, the greatest jump in the use of a particular construction can be observed with the pattern *get*-ProPers-*to*-V, which more than doubles in the group of adults. This is also the only construction where the early groups prefer to use the nominal in the position of the causee as opposed to person pronouns, which they elsewhere almost always prefer. This is not the case with adult speech, but even in adult speech, the frequency of use between the *get*-ProPers-*to*-V and *get*-N/MOD-N-*to*-V comes closer than in other patterns in terms of pronoun-nominal competition. Although it is not completely clear why this might be the case, it surely corroborates the claim that sufficient input frequency is required to 'tip the scales' when it comes to competing patterns in children's language.[54]

The distribution of frequencies when it comes to causative constructions shares some main tendencies observed (albeit not all of them) in both ditransitive and relative constructions across different age groups. The frequency of the overall use of periphrastic causative as in the child-oriented adult speech is almost 2 times greater than in the 0-3 age group. As expected, the frequency at which they produce constructions requiring a greater level of syntactic complexity was lowest in the 0-3 age group, whereas when it comes to 4-6 and adults, the frequencies are approximately the same; the data does indicate that children indeed produce the periphrastic causatives by the age of four, but the discrepancies in frequencies suggest that the learning curve might not be complete. However, the main thing to be observed here is that the overall distribution of different periphrastic causatives demonstrates a striking resemblance between all three studied age groups. The patterns that seem to be most frequent in the adult input are also the ones that are most frequent in early speech, both in 0-3 and 4-6 age groups. The same goes for patterns where the input frequency was at its lowest, that is, the children in any of the two younger age groups did not demonstrate any need for a particular pattern that was underrepresented in the input.

[54] I have also observed the nominal bias towards the infinitival complement in adult speech on the spoken transcripts from the BNC corpora with the causative *help* construction, which might suggest that the lexical differences are for some reason triggered by the syntactic environment on its own (Proroković, in preparation).

Table 16. Distribution of causative constructions across age groups (relative frequencies normalised per million tokens and rounded to the closest unit)

	0-3	**4-6**	**18 +**
***get*-ProDem-*to*-V**	0	2	1
***get*-N/MOD-N-*to*-V**	8	11	15
***get*-ProPers-*to*-V**	5	7	18
***have*-ProDem-V**	1	0	0
***have*-N/MOD-N-V**	1	7	9
***have*-ProPers-V**	1	6	12
***help*-N/MOD-N-V**	5	13	25
***help*-N/MOD-N-*to*-V**	0	0	2
***help*-ProPers-V**	71	71	96
***help*-ProPers-to-V**	1	5	4
***let*-ProDem-V**	0	0	0
***let*-N/MOD-N-V**	23	45	75
***let*-ProPers-V**	374	633	550
***make*-ProDem-V**	1	0	1
***make*-N/MOD-N-V**	16	50	35
***make*-ProPers-V**	75	132	135
***TELL*-NP-*to*-V**[55]	12	82	82
***CAUSE*-NP-*to*-V**[56]	0	0	2
***ALLOW*-NP-*to*-V**[57]	0	0	1
***TALK*-NP-*into*-V**[58]	0	0	0
TOTAL	594	1065	1065

Table 17. Examples of the produced periphrastic causative constructions in the CHILDES corpora (patterns from Table 16 indicated in square brackets)

Construction pattern	**Example**	**Corpus**	**Age**
***get*-ProDem-*to*-V**	*Oh you want to [get some of those to come]?*	WARREN	6;2
***get*-N/MOD-N-*to*-V**	*Um xxx I''m gonna [get the cow to drink some milk].*	BLOOM	2;8;12
get-ProPers-*to*-V	*And my can [get it to go for towing the blue towing in the the blue car].*	NELSON	2;0;1
***have*-ProDem-V**	*Yeah now we need to lock it we can''t [have that happen].*	GARVEY	3;6

[55] ***TELL*** category included the following verbs (as search queries): *order, command, appoint, direct, require, request, demand, dictate, ordain, convince, motivate, urge, ask, beg, bid, tell, instruct, authorize, empower, entitle, endorse, sanction, support, insist, talk.*

[56] ***CAUSE*** category included the following verbs (as search queries): *cause, force, compel, coerce, lead, oblige, drag, drive, press, pressure, blackmail, strong-arm*.

[57] ***ALLOW*** category included the following verbs (as search queries): *allow, permit, enable, grant, license, approve*.

[58] ***TALK*** category included the following verbs (as search queries): *order, command, appoint, direct, require, request, demand, dictate, ordain, convince, motivate, urge, ask, beg, bid, tell, instruct, authorize, empower, entitle, endorse, sanction, support, insist, talk.*

***have*-N/MOD-N-V**	*And then he drove away and then the car [had its wheels blow up] and then he got he flew back and then he xxx.*	WEIST	3;3;9
***have*-ProPers-V**	*Have you wrap me that way.*	WEIST	3;3;9
***help*-N/MOD-N-V**	*Help baby take coat off.*	MCCUNE	1;8
***help*-N/MOD-N-*to*-V**	*You were [helping Granddad to water his courgettes].*	THOMAS	18+
***help*-ProPers-V**	*Do you want to [help me do my Noddy jigsaw]?*	BELFAST	3;7;13
***help*-ProPers-*to*-V**	*Uh [help us to imagine]!*	HSLLD	4;9;23
***let*-ProDem-V**	*Peter couldn''t [let that happen].*	HSLLD	18+
***let*-N/MOD-N-V**	*Slide [let those other people come up].*	VANHOUTEN	3;3;28
***let*-ProPers-V**	*Huh [let me kiss it].*	DEMETRAS	2;7;2
***make*-ProDem-V**	*We [make them all stand up].*	SUPPES	2;10;13
***make*-N/MOD-N-V**	*Can can you [make mummy get some breakfast]?*	LARA	2;10;15
***make*-ProPers-V**	*For [making him catch the dog].*	WOLFHEMP	6;5
***TELL*-NP-to-V**	*Mommy I [might ask you to paint a tulip] because that''s one thing I''m not good at.*	ERVINTRIPP	5;8;11
***CAUSE*-NP-to-V**	*And [forcing someone to eat it] tulips.*	ERVINTRIPP	18+
***ALLOW*-NP-to-V**	*(…) they weren''t letting you they weren''t [allowing you to treat yourself] by relaxing the muscles.*	SNOW	18+
***TALK*-NP-into-V**	*You''re not [talking me into letting you have that kitty here]?*	HSLLD	18+

Furthermore, the Spearman's rank correlation indicates high values when it comes to frequencies of the observed patterns within the causative construction (N=12).[59] The correlation is signficant on the 0.01 level and highest among the 0-3 age group and the adult group, and nearly the same value is found between 0-4 age group and the adults (see Table 18). Again, there is little to no difference between the frequency-based ranking of particular periphrastic causatives, which leads us to assume the same degree of correspondence in terms of preferred patterns across all three age groups.

Table 18. Spearman's rank correlation of the observed causative patterns across age groups

	0-3	**4-6**	**18+**
0-3	**1**	**0.923****	**0.958****
4-6		**1**	**0.951****
18+			**1**

**ρ=0,591 p=0.05 **ρ=0,777 p=0.01*

[59] Note that the Spearman's rank correlation was conducted for the first 12 patterns because some were simply not found in the corpus for the youngest age groups, meaning that they could not be ranked amongst themselves.

Although there is a quite a bit of similarity across age groups in terms of distribution and frequency of usage, there are some interesting discrepancies can be more easily observed when ratios are taken into account. For instance, *help*-ProPers-V is one of the more frequent patterns found in adult speech, so it is not surprising that the same goes for younger age groups. However, the observed volume of production in younger age groups comes very close to adult speech (which means that, comparatively speaking, children produce it more than other patterns when contrasted with adults). This could be foreseen in 4-6 age groups, especially considering that the overall production of periphrastic causatives appears to be the same between them and adults. If anything, given the same overall volume of periphrastic causatives in the two groups, one could argue that the constructions somewhat drops in terms of expected frequency in 4-6 age group. However, given the rest of tendencies between 0-3 age groups and 4-6 age groups, which typically reflect a rise in the use of various causatives by a significant margin, it may be interesting to consider why the same rise is not exhibited in *help*-ProPers-V, i.e. why the same construction is so frequent from the onset. When compared to other causative patterns, this is the only one which does not rise considerably once the children turn 4 years of age. Well, one of the reasons that may explain this anomaly stems from the fact that the youngest age groups simply need to use the pattern at question more often. In other words, the situational context that the youngest children find themselves in, or the requirements they experience that necessarily need to be fulfilled by their interlocutors, lead them to use the construction in question. For example, it is certainly expected that children constantly need some sort of help with various tasks at hand, at which point they resort to constructions such as *are you going to help me make this* (BELFAST, 2;7;26) or *help me move it out of the way* (BLOOM, 2;10;19) etc. To further support this, the analysis revealed that out of the total 301 instances observed in the corpora (of the patterns *help*-ProPers-V and *help*-ProPers-*to*-V), 240 of them contained *me* as the causee pronoun (79.7%). Although the construction was not as pronouncedly represented in the adult speech, it was represented sufficiently for children to make early generalisations and inferences about the flexible use of that particular pattern (out of the total 1874, there were 936 instances with me in the causee position – or 49.9% in total).

The second pattern where the tendencies differ across age groups more than they do with some other patterns is *TELL*-NP-*to*-V, which included a variety of verbs that have a causative implication to them. The analysis of the said pattern, along with the other two non-

prototypical periphrastic causatives considered (*CAUSE*-NP-*to*-V and *ALLOW*-NP-*to*-V), located only three verbs in the 0-3 and 4-6 age groups (Table 19). In other words, despite the fact that a large variety of verbs with causative implications were included in the analysis (43 of them; see Footnote 55, Footnote 56, and Footnote 57), only three types of instantiations were observed in 0-3 and 4-6 age groups. In the 0-3 and 4-6 age groups, we find the verbs *tell, ask* and *allow*, whereas in the adult group we find other verbs such as *force, cause, order, press, beg, led, convince, pressed, drive, urge, talk, enable, drag, get, demand, request* and *command*. Not only did the extraction method yield just three different verbs in the causative pattern in the younger age groups, but the verbs *allow* (in the 0-3 age group) and *command* (in the 4-6 age group) were observed only once.

Table 19. Observed frequencies of particular verbs in the pattern V NP to V (excluding get and help)

0-3			**4-6**			**18+**		
tell	36	72%	*tell*	111	82.84%	*tell*	931	58.33%
ask	13	26%	*ask*	22	16.42%	*ask*	578	36.22%
allow	1	2%	*command*	1	0.75%	*allow*	24	1.5%
Total		100%			100%			96.05%

The data in Table 19 clearly shows another similarity across age groups within the V-NP-*to*-V pattern. Regardless of the fact that some instances of other verbs were found in the adult speech, it is clear that two verbs represent the majority of all periphrastic causatives when it comes to child-directed speech. But what seems to be even more interesting is that the same two are the most frequent ones in early speech. The results can be interpreted in the usage-based framework, as they also seem to favour the item-based tendencies in early speech. The most frequent verbs that the children hear are also the ones which completely rely on in their early speech. This is not to say that children do not necessarily understand or use the rest, especially as they grow older and reach 5 or 6 years of age, but it is indicative of the way in which they express themselves early on. Moreover, aside from the fact that *tell* and *ask* are most frequently found in the input, it is important to point out that they are also the ones which are semantically most flexible. We could easily look at the two as 'umbrella terms' for many others that have been included in the analysis, which tend to be more restricted in their meaning, entailing much more nuance when in terms of goals, participants in the action, manner etc. The fact that they are semantically flexible reflects on the speech of children, as well as that of adults; however,

it is safe to assume that adults may be even more lexically diverse (see Section 4.2.5) in their periphrastic causatives had the analysis included non-child directed speech. The distribution of the verbs only corroborates the usage-based claims of first verbs learned in constructions being the most frequent ones and relatively easy to understand in comparison to others; although *tell* is comparatively speaking relatively easier than *request* because it is less nuanced, it is still more abstract when compared to *make*, and the differences in the frequency of use between the two in the 0-3 age group reflect this fact.

As far as the competition between the infinitival periphrastic causatives and bare-infinitive ones is concerned, the results indicate that all age groups clearly favour the latter form (Table 20); Chi-square test shows that the difference in the rates is significant, χ^2 (2, N = 2) = 25.2585, $p < 0.05$. Clearly, the distinction between the two and the corresponding results need to be interpreted carefully, since the two causative patterns take different verbs, which means that the disproportion is directly affected by the selection of verbs and the variety of situations in which their meaning is suitable or applicable. Even if we remove the *let*-NP-V pattern from this analysis (see Table 20), which does not constitute a canonical periphrastic causative and arguably has the meaning that is most distant from the rest of the verbs included in the overview, the pattern V-NP-V still remains dominant when contrasted with V-NP-*to*-V. Furthermore, the disproportion between the two patterns is the greatest in the 0-3 age group; while the ratios are approximately 1:9 in the 4-6 age group and adult speech, there seems to be at least 20 V-NP-V on 1 V-NP-*to*-V in 0-3 age group. The disproportion remains similar even when we remove *let* causatives from the equation. Such data supports the proposition that the early language tends to be saturated around the most frequent items it received in the input. Similarly, the most frequent constructions in the input could also be even more frequent in the earliest speech, causing the disproportion to be even higher than usually expected. Although the difference between the two does not directly concern the lexical choices in causatives, much of the observed disproportion is related to the fact that children produced much less verbs in the 'tell category'.

Table 20. Distribution of causative constructions across age groups (relative frequencies normalised per million tokens; some patterns are merged together)

	0-3	**4-6**	**18 +**
V-NP-*to*-V	26	108	126
V-NP-V	568	957	939
V-NP-V (without *let* causative)	171	279	314

Furthermore, the only verb that takes both variants is *help*, and it is indicative that the variant with the bare-infinitive tends to be the preferred form. One possibility is that the majority of other causative constructions tend to take the form employing the bare-infinitive, which then affects the choice of the pattern *help*-NP-V as opposed to *help*-NP-*to*-V. In other words, speakers (especially children) may be affected in their choice of the preferred pattern in the causative *help* construction simply by observing other most frequent causative patterns and generalise in case of the causative *help*. Another possibility is that the causative pattern which employs the 'full infinitive' is simply less economical than the other, which should in turn affect the frequency of the choice.[60] An interesting take on the issue with help alternation has been provided by Lohman (2011) who exploited Rohdenburg's Complexity Principle (1996, p. 151), i.e. the Principle predicts that the language user in a context of high complexity will use more explicit full infinitives, while in the case of low complexity the more economical bare infinitive would be the preferred option. In other words, the speaker is more likely to "sacrifice" the economic aspect of the produced construction for the sake of clarity (if they believe that the more explicit variant would bring about less confusion). The fact that there is a less pronounced discrepancy in the production ratio as the age increases (in 4-6 age group and adults it is approx. 9 to 1 in favour of bare infinitives, whereas in 0-3 age group it is 20 to 1) might suggest a greater awareness of the language use on part of the older age groups even in the conversational setting typical of CHILDES subcorpora, but a more robust argument for this could only be made with further comparative analysis of other spoken corpus data.

[60] A similar parallel phenomenon occurs with semi-modal verbs such as *dare* or *need*, though the alternation tends to be reverse when it comes to the structural alternation in the function of time. In a corpus-based approach Taeymans (2004) found that the semi-modal *need to* has become the dominant form in the contemporary English as opposed to its Early Modern English counterpart *need + bare infinitive* (retained in language use predominantly with negation *need not*), which she considers to be a "breach" of the unidirectional movement along the grammaticalization cline. Though her data did not find such conclusive direction with *dare*, the evidence pertaining to *need* that goes against the more economic bare-infinitive variant could indicate a semantically motivated shift in grammatical structure.

4.2.3. Lexical arrangement and overlap

In total, the verbs heading periphrastic causative constructions show a considerable overlap in the frequency of usage across all three age groups (see Table 21). There is a significant inter-group difference to which they differ in terms of overall production rates, χ^2 (8, N = 5) = 24.3719, $p < 0.05$, but there is also a clear overlap in view of the relative proportions of use. Not only does the *let* periphrastic causative make up the majority of their production, but the rest of the verbs share a similar position in the use hierarchy. However, it is important to observe a greater concentration of periphrastic causatives around the most frequent verbs in the 0-3 age group as opposed to others (97.95% in relation to 92% approx. for 4-6 and adults). Also, proportion-wise, there is a greater use of *let* construction in the 0-3 age group than in others, indicating a more conspicuous item-based trend in the production.

Table 21. Main verbs in the periphrastic causatives

	0-3		**4-6**		**18 +**	
have	3	0.56%	13	1.26%	22	2.07%
get	13	2.13%	20	1.89%	34	3.23%
help	78	13.06%	89	8.37%	127	11.91%
make	92	15.47%	182	17.09%	171	16.06%
let	397	66.73%	678	63.65%	625	58.71%
TOTAL	582	97.95%	983	92.26%	979	91.97%

An important observation regarding the verbs concerns the usage-based perspective and the relevance of cognitive capacity to comprehend more abstract concepts. For instance, if we approach the verbs *have, get, help, make* and *let* on their own, outside of the context a construction tends to provide, we might argue that some are more complex than others – in the sense that they differ in their level of abstraction. For example, the verb *make* tends to be rather straightforward and often entails an action observable to children in the physical world, whether in transitive (*make a cake*) or ditransitive (*make me a cake*). Things start to get slightly more complicated when we enter the realm of periphrastic causatives (*make him bake a cake*) where the verb *make* no longer denotes an action which is necessarily observable in itself, as the 'act of forcing' somebody to do something might not be as explicit or evident to the child as 'making a cake' might be. In this sense, we could expect that the child's previous knowledge of the verb *make* and the way in which it is used (e.g. the transitive *make*) might conflate with their

understating of the new, more abstract construction (e.g. the causative *make*) and make the process of acquisition more difficult. On the other hand, we might assume that the previous knowledge helped them in a 'pathbreaking' manner and made the new meaning more accessible, thus making it more easier. This is especially interesting if we consider the underrepresentation of the causative *tell* in the 0-3 age group and compare it to *make*. There is no particular reason why 'make' should be more frequent in the causative pattern than the verb *have* given that they share the same underlying idea. First, we may argue that the verb 'have' is a more conceptually difficult verb from the start, as it usually entails 'possession' (e.g. *I have a dog*),[61] and that much of the children's cognitive capacity is first directed towards unlocking the 'usual' meaning and use of the verb *have*. On the other hand, the most obvious explanation is clearly that the periphrastic causative *make* is much more frequently found in the input, which makes it easier for children to understand, reproduce and generalise on the variety of situations in which the verb can be used.

The following step in the study of causatives and their differences across age groups looked at the most frequent nouns that take the role of the causee in the periphrastic causative pattern V-NP-V (Table 22). While the acquisition of the construction tends to be primarily anchored in verbs that head the constructions in question, it may be beneficial to expand the same logic to nominal elements, i.e. a high degree of resemblance in terms of nominal elements in a construction such as the causative one may prove to be a true testament to the 'item-based' tendencies and the partially 'imitative' character of early language. In order to see whether the acquisition of causative construction anchors itself around the most frequent nouns in the input, it is useful to look at the most frequent lexical choices made by the studied age groups.

[61] The causative *have* in many ways seems to be a result of a grammaticalization process, whereby the verb *have* goes through a resultative interpretation of the possession: *I have a dog → I have a dog (which is) killed → I have killed a dog* (cf. Ziegler, 2004; Kouteva, 2019, pp. 342-343)

Table 22. Frequent nouns (causee) in the causative pattern V-NP-V[62]

Frequency order	0-3	4-6	18+
1	proper names	proper names	*mommy* (and variations)
2	*people*	*people*	proper names
3	*mommy/mummy*	*daddy/dad*	*daddy/dad*
4	*car(s)*	*mommy/mom*	*car(s)*
5	*daddy*	*car(s)*	*thing(s)*
6	*horse/horsie*	*(no/any/some/every)body*	*man/men*
7	*guy(s)*	*things*	*people*
8	*thing(s)*	*guy(s)*	*train(s)*
9	*baby*	*water*	*baby(ies)*
10	*pig(s)/piglets*	*man*	*teeth/tooth*
11	*animals*	*light*	*eyes*
12	*dolly*	*feet*	*hand(s)*
13	*lion*	*dinosaur(s)*	*animals*
14	*children*	*hand(s)*	*guy(s)*
15	*doctor*	*(any/every)thing*	*boy(s)*

One of the questions that is constantly re-examined in language acquisition studies is whether children initially learn particular constructions as a whole, repeating and producing the lexemes that are most frequently found in the input. These suppositions have been somewhat confirmed and corroborated in terms of verbs that head the constructions (see Ibbotson & Tomasello, 2009; Hilpert & Diessel, 2017), but the research is less conclusive when it comes to nouns which do not occupy a central role in constructions that are being acquired. Naturally, it would not be sound to expect the same amount of similarity with nouns as one does with verbs, primarily because the constructions such as the causative one allow a far greater variability in the slots filled with noun phrases. In other words, a huge similarity across age groups in terms of nouns found in their causative constructions is less likely by default.

The first thing that becomes apparent in the data is the fact that proper names take the first position in terms of item frequency in the place of the causee in the age groups 0-3 and 4-6, while in the adult group, they take the second position. Of course, the main reason for this is the fact that a number of different names were lumped together into this category, so it is

[62] This analysis includes the causative verbs *make*, *have*, *help* and *let* that occur in the pattern 'causative verb-causee-base verb'; in 4-6 age groups the second most frequent causee was actually *balloons* but it was removed from the final analysis because it was included in the Hicks corpus which elicited early speech with the help of storybooks, and in this case the story highly motivated the speakers to use *balloon* in the causative manner, thus corrupting the data for the purposes of this comparative analysis.

prudent to observe these as a category of its own (see the later analysis of animacy in causatives). As far as specific nouns are concerned, it can be observed that the first few choices in all age groups revolve around variations of the nouns *mom* and *dad*, but also include others such as *people, cars, things* etc. It is not surprising that the variations of mom and dad are among the most frequent nouns that are found in all age groups in the position of the causee. Much of the early language is directed towards parents, at least as observed in the CHILDES corpora. The same goes for parents or investigators who often refer to parents in third person singular. It is also interesting to note that *people* is actually the most frequent noun in the position of the causee in the early language (if we take out the *proper names* from the list), as opposed to adults where it takes the 7th (or the 6th) position. Certainly, the similarity across age groups in terms of nouns in causative constructions can primarily be explained by the fact that these are the nouns which are the most frequent ones in language in general, especially in terms of animate beings.

Interestingly, *car(s)* tend to be very frequent in the causative pattern across all age groups. While the animate beings are restricted to considerably fewer words that denote them in general, which is why it is less surprising to see them repeated across all age groups in the same pattern, inanimate nouns are a much wider category, thus reducing the likelihood of such resemblance in frequency. The fact that this noun is one of the most frequent ones agrees with usage-based claims which incorporate the role of cognition and understanding into learning that relies on the most frequent items in the input. Although the notion of *car* in itself represents (or may be categorized as) something inanimate, there may be less semantic constraints on it than there would be on some other inanimates such as *stone*.[63] For example, it can take a wider range of verbs concerning movement and operating complex machinery (e.g. *go/drive/drift/turn*/*stop/run* etc.),[64] thus making its usage in a construction such as the causative

[63] Because of their specific properties, 'vehicles' have sometimes been separated from other inanimate objects, especially in the context of language acquisition and development (see Graham and Poulin-Dubois, 1999; Arterberry and Bornstein, 2001; Opfer and Gelman, 2011; Rostad et al., 2012; Baker et al., 2014 etc.), with many of them having concerned themselves with animal-vehicle distinction when some of their parts have been purposefully altered and replaced, such as wheels and legs (Mandler et al., 1991; Rakison and Butterworth, 1998; Rostad et al., 2012 etc.). Rostad et al. (2012) have looked into the ways in which infants categorize specific objects (partonomic vs. taxonomic), and found that the items modified for the purposes of their experiments were primarily perceived through animate/inanimate properties that were arguably less obvious than other perceptual features which would have steered the categorization into another direction.

[64] Interestingly, the study by Baker et al. (2014) found that infants as young as 10 months old demonstrated the ability to limit the motion path of jumping to animate beings while excluding vehicles, which in turn they interpreted as a support for the hypothesis that infants early in development associate motion paths with distinctions between animate and inanimate entities.

one open to a greater number of schemas with different caused event verbs (e.g. *make the car run* vs. *make the stone run*). The fact that all age groups produce it so frequently in the causative construction tells us that children start to grasp this concept very early, both the intent expressed in the causative construction and potentially the difference between *cars* and other inanimate nouns (e.g. *stone*).

Other frequent animate nouns in 0-3 age group include animals which usually refer to toys or animals in storybooks or other materials used to stimulate/elicit speech. In other words, the frequent causees in the 0-3 age group tend to be concrete objects, those that are observable or accessible to children. Most of these are not found in the input among the most frequent ones, which indicates that children utilize the function of causatives for expressing activities and notions relevant to them in immediate surroundings (*my car* in *make my car go*; BOHANNON, 3;0), whereas adults apply causatives more frequently on situationally less specific concepts (e.g. *eyes* in *sometimes onions make your eyes cry*; FLUSBERG; 18+). Moreover, with the progression of age, in the 4-6 age group the most frequent ones become more abstract (for example, *light*, *water*, *anybody*, *anything* and the like). This indicates both a linguistic and a cognitive development; (1) linguistic in the sense of the ability to generalise on the use of more abstract concepts in causative constructions and to apply them productively and flexibly, and (2) cognitive in the sense of understanding more complex phenomena, but more importantly, understanding which conditions they ought to satisfy to fill the 'causee slot' in the causative construction.

When looking at the secondary verbs (verbs heading the caused event) in the periphrastic causative constructions, there again seems to be a considerable overlap in the most frequent choices for all age groups. Despite the fact that the secondary verb slot in the periphrastic causative construction is arguably constraint free semantics-wise (it is more or less possible to fill the slot with any verb), the cross age comparison indicates a resemblance in different age group preferences in using particular verbs as opposed to others.[65] For example, the verb *go* is the most frequent choice in all age groups, and the verb *do* also ranks highly for all three, while the verbs *get*, *come*, *feel*, *disappear* appear among the most frequent choices in

[65] It is true that some verbs make better "candidates" for filling the slot, as they tend to depict an action which can indeed be triggered by the actions of another person (causer). Action verbs may be preferred as opposed to stative verbs (*make someone go* vs. *make someone be*), though not exclusively (for example, consider the frequently used *make someone think* or *make someone feel*).

all age groups. Interestingly, the verb *get* is among the top three choices for the 0-3 and 4-6 age groups, whereas in the adult group, though frequent, ranks considerably lower. This may be anchored in situation-specific or participant-specific factors which drive one item to be produced as opposed to others. For example, children may use *get* more frequently because they tend to be in need of something more often than adults (e.g. *can you make mummy get some breakfast*; LARA, 2;10;15), whereas *feel* may rank more highly in adult speech due to their appeal to emotions in children (e.g. *does it make all your muscles feel better*; FLUSBERG, 18+). It is difficult to assess just how much of an impact the input bears in relation to other pragmatic aspects of language production (and reproduction), but the overproduction of the same items that most frequently occupied certain slots in the input indicates item-based patterns that go beyond items heading certain phrases (such as ditransitive or causative verbs).

Table 23. Secondary (caused event) verb in the V-NP-V pattern frequency ranking across age groups

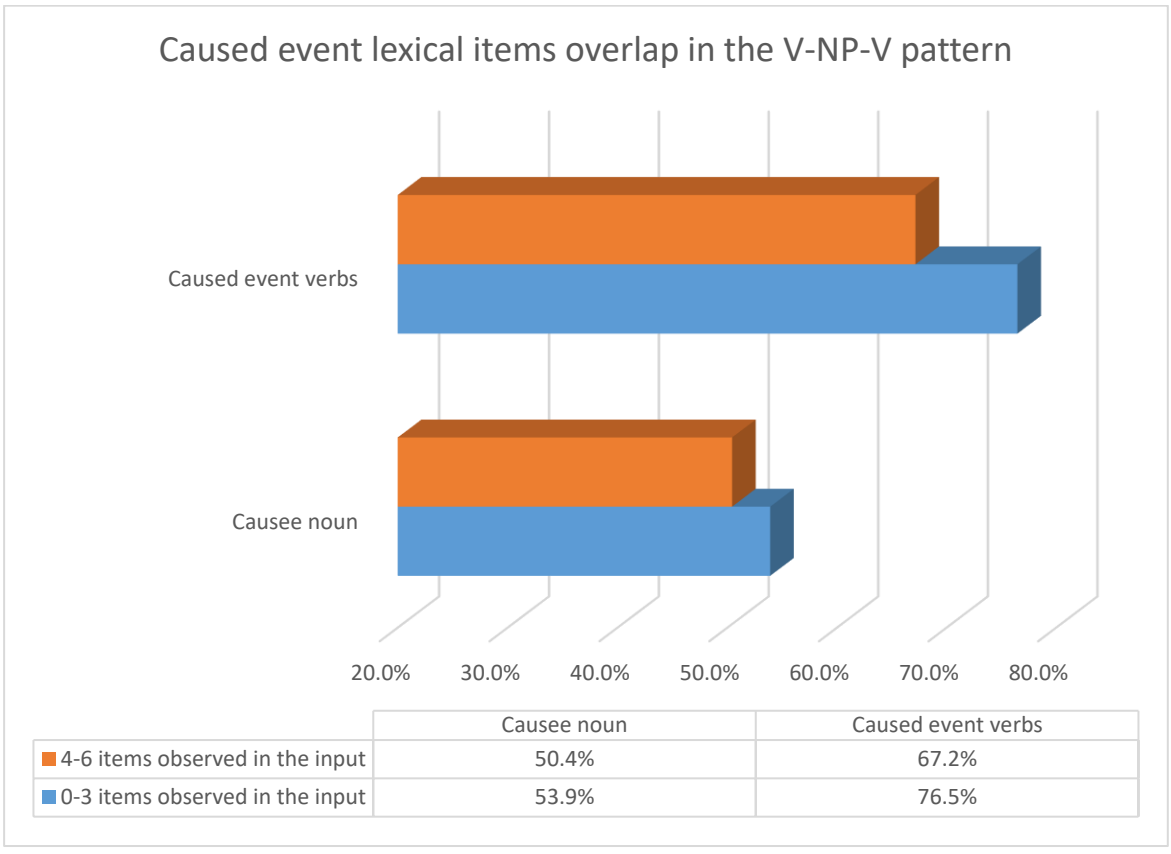
Rank	0-3	4-6	18+
1	*go*	*go*	*go*
2	*do*	*get*	*do*
3	*get*	*do*	*feel*
4	*come*	*grow*	*sit*
5	*stand*	*play*	*look*
6	*see*	*feel*	*come*
7	*feel*	*come*	*stand*
8	*sit*	*fly*	*get*
9	*eat*	*clean*	*put*
10	*write*	*cool*	*make*
11	*grow*	*disappear*	*fall*
12	*say*	*fall*	*disappear*
13	*disappear*	*stand*	*eat*

Of course, there are other important mechanisms that might be at play besides the pragmatic element of their speech and the relevance of the input. The ranking of certain verbs can be affected by the vocabulary deficiencies in the youngest age group. One of the things that we may notice is the high frequency of the verb *see* in the 0-3 age group, and one of the reasons why this might be the case is their inability to retrieve a more straightforward lexical item to express the intended meaning as opposed to using the periphrastic variant. In other words, construction such as *make him see* is another way of saying *show him*, and a similar relationship

(to a certain degree) is captured with *help* (both have been observed concomitantly with *see* in the 0-3 age group). The compensation of lexical inaptitude with syntactic prowess is a reflection of the fact that the development of one's vocabulary is a lifetime-long process, i.e. the younger the children are, the higher the chances of them resorting to periphrastic alternatives if the more straightforward lexical ones are not retrievable at the given moment.

A more detailed insight into the correspondence of verbs within the periphrastic causative construction is better illustrated in the level of overlap between the input (adult language) and the two age groups of children. The exact degree to which there is a similarity in using certain lexical items in the specific slots within the causative construction, is more clearly visible in the cross-group differences in the proportions of items that were also found in the input (see Figure 8). When it comes to both caused event verbs and causee nouns, the proportion of items that were also observed in the input is greater for the 0-3 age group than for the 4-6 age group.

Figure 8. The proportion of lexical items observed in the input in the periphrastic V-NP-V pattern



	Causee noun	Caused event verbs
4-6 items observed in the input	50.4%	67.2%
0-3 items observed in the input	53.9%	76.5%

It is not particularly surprising that both age groups tend to overlap to a considerably greater extent when it comes to the verb slot, than when it comes to the noun slot. The primary reason for this is probably vocabulary-related while the other may be of pragmatic origin. In other words, the range of lexical items is naturally smaller for verbs than for nouns, while the second reason is probably the difference between children and adults when it comes to their choice of the referent; i.e. the production of periphrastic causative constructions entails that the speaker choses a particular causee which is to be forced into doing something, and the focus of children and their interlocutors is often oriented on different actors or entities within the experimental setting. Nevertheless, the results clearly indicate that there are more lexical items that have been observed in the adult speech produced by the 0-3 age group than by the 4-6 age group, which suggest a more restrictive character of the earliest speech as predicted by the first hypothesis (see hypothesis 1.3).

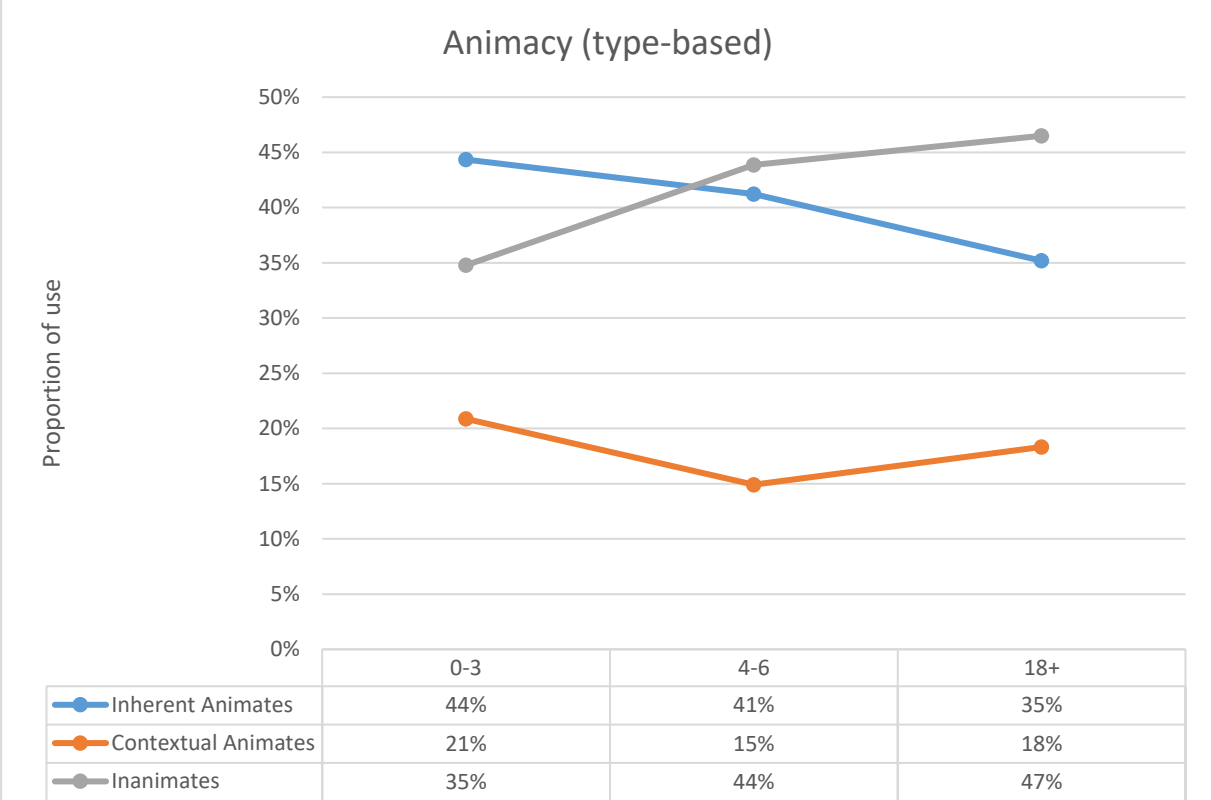
4.2.4. Animacy

As far as the role of causee is concerned in the V-NP-V pattern, a look at the animacy may provide a new perspective on the language production in relation to age differences and the character of the nouns used in the causee slot. This element takes the central position in the causative construction and is crucial for the interpretation of the entire construction. For example, it may take both inanimate and animate forms, although it should arguably favour the animate ones given that they are expected to carry out an additional action, i.e. given that the causee tends to be the agent of the secondary action, it should more likely take on the animate form. Although this is a bold claim, a look at the most frequent caused event verbs reveals a potentially verb-induced constraints on the causee slot (see Table 23); i.e. the vast majority of them (across all age groups) are verbs that tend to semantically entail a relationship where the subject takes on the sentient form (e.g. *feel, sit, look, come, stand, get, make, eat, write, say* etc.). In other words, if we assume the greater likelihood of the causee taking on the role of the agent (or a sentient patient), we might also assume a greater likelihood of said causee to take on the animate form. This type of comparative analysis may shed additional information on the perception of animacy in terms of construction processing on the part of children; we may partly draw inferences on the competitiveness between the mechanisms of understanding/generalising (e.g. the semantic implication of the most frequent caused event verbs observed in the

periphrastic causatives should drive children to employ animate agents in the position of the causee) and input frequency (e.g. the most dominant pattern in the input should affect the children's production regardless of its suitability).

The results on the three categories of animates (inherent animates, contextual animates and inanimates) reveal where the shift in animacy actually occurs across the age groups in the V-NP-V pattern. It has already been shown that the use of inanimate nouns in the position of the causee rises with the progression of age; it rises considerably between the 0-3 age group and the 4-6 age group, and it continues to slightly rise in adult language. Naturally, the rise of inanimates comes at the expense of other categories of animacy and the Figure 9 illustrates the trends across age groups.

Figure 9. Animacy of the causee in periphrastic causatives



The rising trend in the use of inanimates that comes with the progression of age is accompanied with the declining trend in the use of inherent animates in the position of the causee. It is the highest in the 0-3 age group (44% of all uses), but drops in the 4-6 age group (41% of all uses) and continues to drop in the adult group (35% of all uses). The portion of different contextual animates employed in the 0-3 age group amounts up to 21% of all uses,

dropping to 15% in the 4-6 age group, and then rising again in adult language. If we go to Table 24, we can clearly see that the drop between the 0-3 age group and the 4-6 age group is in the use of animate toys in the position of the causee. This is not completely surprising, given that we can expect the youngest category to be oriented towards toys more than slightly older children. With the rise in cognitive power and the ability to handle more complex phenomena in terms of conceptualization, the children in the 4-6 age group may employ causatives anchored to their immediate surroundings less frequently (e.g. *cause this makes the light come when I'm talking*; HALL, 4;6). On the other hand, when it comes to adults, the examined speech was mostly child-directed which is why they may have resorted to animate toys more frequently when requesting of their young interlocutors to perform a certain action (e.g. *make the fish bite the truck*; SACHS, 18+).

Table 24. Animacy of the causee

	0-3	**4-6**	**18+**
INAN	35%	44%	47%
ANc	23%	23%	16%
ANp	22%	18%	19%
ANt	14%	8%	13%
ANb	7%	7%	6%

When the animate causees are split into 4 categories based on the degree and context, inanimate things appear to be the most frequent ones across all age groups in the position of the causee. Naturally, if the rest of the categories of animacy were lumped together into one group, they would be more frequent than inanimates, but considering that these do not constitute a coherent and homogenous group, it seems best to treat them separately in this analysis. Although the overall patterns in terms of (in)animacy of the causee largely coincide across age groups, there seem to be some smaller discrepancies which may be telling for the differences in the use of causatives. The first thing to observe is that the adult group employs inanimate items more frequently than the younger age groups, but especially when compared to the 0-3 age group. Indeed, inanimates are fairly frequent in all age groups, but the saturation around inanimates tends to be greater in the adult group. One possibility is that the rise of lexical richness in the adult group results in the greater portion of inanimate items—if we presume that there is a greater variety of nouns that are inanimate, that is, that with the greater number of diverse nouns there is also a greater potential for the production of inanimates. However, the

amount of inanimate items in the position of the causee may suggest an additional aspect of the causative construction that becomes more prominent as the children grow older. An inanimate item in the position of the causee may suggest that the construction moves away from the prototypical role of the agent in the position of the causee (e.g. *have them go this way*; KUCZAJ, 3;4;8) towards the role of the patient (e.g. *make my milk disappear*; BLOOM, 2;7;13). To be fair, many of the nouns labelled as inanimate adopt animate characteristics and perform the roles of the agent in some sense (e.g. *make the wheel go into the barrel*; BLOOM, 2;8;12), but it seems legitimate to assume that they require a greater degree of abstraction and conceptualization than the typical animate agents.

Whether the prototypical periphrastic causative construction employs the causee as agent rather than the patient of the secondary action is less of an issue here; in the context of this research, it is more important to observe the fluid function of the causative construction and how it affects its development and acquisition. In other words, the input children receive permits the use of inanimate items in the position of the causee to a surprisingly high degree, which should in turn result in similar numbers when it comes to the early language output. Indeed, the portion of inanimate nouns in the position of the causee is fairly frequent in the 0-3 age group (35% of all uses), but still considerably less than in 4-6 age group (44% of all uses) or the adult group (47% of all uses). This may indicate that in the competition between the cognitive mechanisms related to understanding (how children initially interpret the causative constructions and their function) and input frequency (what children hear the most), the limits of early cognition may somewhat skew the results, resulting in a preference towards animate beings in the position of the causee (which represents less of a conceptual challenge for children). Distributional learning plays a role in the sense that children learn to expect animate beings in the position of the agent across other transitive constructions, which may change and become more flexible in the causative construction (e.g. *make the bed stand up and put the little person on it*; BATES, 18+). When confronted with first causatives, for the sake of easier processing, children may focus on the ones which take animate beings in the position of the cause as this would not contradict the expectations they developed through the mechanisms of distributional learning. Finally, this interpretation of the results agrees well with Tomasello's hypothesis about imitative learning which incorporates various cognitive mechanisms such as distributional learning, i.e. both input frequency and children's ability to interpret the input and

formulate generalisations have to be taken into consideration for the complete account of early language development.

4.2.5. Lexical richness

In order to get a better grasp on the length and complexity of sentences containing periphrastic causatives, we may look at the data pertaining to MLU (Table 25).[66] The difference in MLUs actually suggests that the most complex (or the longest) sentences containing causatives are found in the 4-6 age group, and not the one of adults. We can now observe that the data differs from the one on ditransitive constructions, at least in terms of which age group produces the most elaborate sentences that contain the construction in question. While the inconsistency in cross-constructional MLU data is described in more detail later (see Section 6.1), we can note that there is a large gap between the 0-3 and the 4-6 age groups, i.e. despite a lot of "noise" and nonce words being recorded in the youngest age groups, the data suggests that the 4-6 age group tends to produce sentences longer by at least 2 words on average when compared to the 0-3 age group.

Table 25. General overview of the sentences containing make-NP-V across age groups

	0-3	**4-6**	**18+**
Token count	3129	3163	29756
Word type count	457	536	1909
Character count (without spaces)	11591	11866	115342
Utterance count	386	298	3204
Open class words	1629	1543	14789
Closed class words	1454	1547	14382
MLU (morpheme-based)	8,59	11,25	9,82

When it comes to lexical density in utterances that centre around the periphrastic causative *make*-NP-V constructions, the numbers are very similar across age groups (see Table 26). Similarly to sentences with ditransitive constructions, the variation in lexical density in those containing causative constructions seems non-existent. The LD values for these

[66] Again, it is important to note that these measures capture the sentences in their entirety, rather than just being restricted to the construction themselves, which means that they might encompass some additional sentence material that does not fall into the causative domain.

utterances revolve around 0.5, meaning that in utterances containing *make*-NP-V causative patterns for each content word there is also one grammatical word on average. The numbers are slightly higher than in utterances containing ditransitive constructions (discussed in the subsequent chapters in more detail), but the point remains that neither of the age groups produces utterances denser in information when compared to the two others. Considering that it centres around two verbs and (at least) one noun, the periphrastic causative construction in itself favours lexical words as opposed to function ones (for example, we might expect higher averages than 0.5). Again, we could expect that lexical density would increase in the younger age groups if they use more isolated forms of causative constructions instead of embedding them in more elaborate utterances, but this does not appear to be the case.

Table 26. Lexical richness for the causative make-NP-V

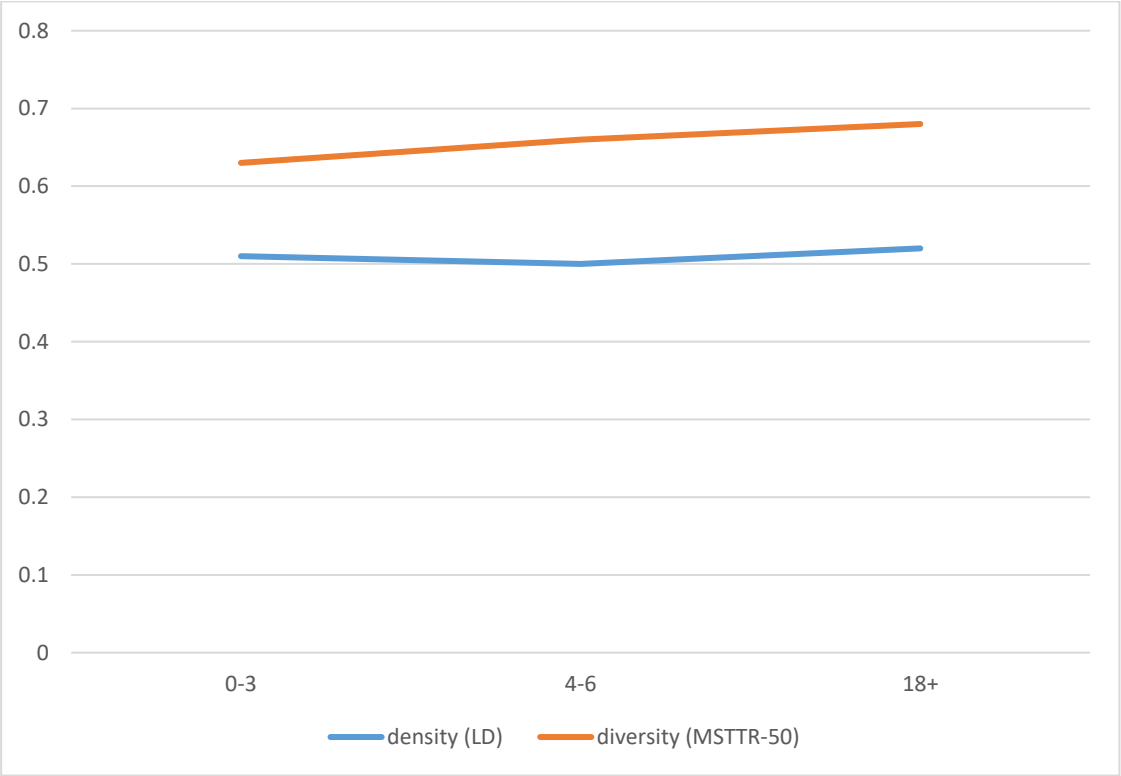
	Lexical density and sophistication			**Lexical diversity: range and variation**				
	LD	LS1	CVS1	NDW	NDW-ER50	MSTTR-50	RTTR	SVV1
0-3	0.51	0.14	0.56	321	34.8	0.63	7.62	10.17
4-6	0.5	0.14	0.8	417	32.9	0.66	8.91	21.48
18+	0.52	0.19	0.86	427	35	0.68	9.8	20.96

When it comes to lexical sophistication (LS1), an increase is only visible between the 4-6 age group and adults. Interestingly, the speakers of all three age groups use words outside of the most frequent ones on less than 2 occasions per 10 words in utterances with causative constructions. While there is an increase in the child-directed language of adults, the question is how much of that has been intentionally simplified in terms of vocabulary. In this context, verb-related sophistication may be more interesting, particularly because the second verb in the construction is open to unconstrained insertion of all sorts of verbs. This index reveals a considerable spike between the 0-3 and the 4-6 age groups already, suggesting that children older than 3 increase the use of less frequent verbs. Additionally, the same trend can be observed even in terms of verb variation (SVV1), which shows a clear spike between the 0-3 and the 4-6 age groups, with little difference between the 4-6 age group and adults. In a sense, the measure of sophistication can be interpreted as an indicator of how much the child relies on the input. As stated before, sophistication is not really a measure of diversity (although they tend to correlate), but a measure of how many 'rare' words a speaker uses. If we assume that the input has a key role in early language development, going beyond the most frequent words in the input suggests a step away from its impact. Sophistication, especially verbal (since verbs head

phrases and are seen as integral in language development) can serve to illuminate a dynamic relationship between the role of input, word frequency and the children's productive language use. If the production of less frequent items can consistently be tracked in its stepwise growth, the argument can serve to further bolster the claims of most frequent items being acquired first, and later guiding the acquisition of more complex (or in this case rarer) words and phrases.

The measures of lexical diversity across age groups also indicate a mild rise in sentences containing causative *make*-NP-V constructions. In terms of MSTTR-50, there is an increase between the 0-3 and 4-6 age groups, as well as between the 4-6 age group and that of adults. The same goes for RTTR, with both values suggesting a slightly greater difference between the two young age groups than between the 4-6 age group and adults. When interpreted together with stable lexical density, we begin to understand that the increase in diversity, regardless of its size, reflects an expansion of utterances with lexical and grammatical items in proportionate numbers, or it shows an increase in the number of word forms without the increase in tokens. For example, if the utterances grew in the number of different lexical items solely, the diversity would increase, but so would density. It is true that the increase in diversity is relatively mild and it is questionable to which degree it match the hypothesized progression in LD. Nevertheless, given that we know that the utterances containing causative constructions grew in size between the 0-3 and 4-6 age groups, but decreased in size between the 4-6 and the adult group, we can speculate that the measures of diversity truly reflect the number of different word forms averaged per the number of tokens.

Figure 10. Density and diversity in the causative make-NP-V



Naturally, the type-token ratios directly relate to the number of different words uttered by each of the age groups. Perhaps the most drastic spike can be observed in the range of lexical items used between the 0-3 and the 4-6 age groups. However, while there is an evident spike in NDW when it comes to utterances with the *make*-NP-V pattern, the averaged NDW of 10 random subsamples (NDW-ER50) tells a different story between the two groups. Given that the latter measure takes into account the number of tokens, instead of providing raw data equalized only according to the number of utterances, the spike in the overall number of different words between the 0-3 and the 4-6 age groups takes on a milder form.

4.3. In sum

- Unlike ditransitive constructions, periphrastic causative constructions were equally present in child-directed speech for both the 4-6 age group and adults, but there was still a significant leap between 0-3 age group and 4-6 in terms of sheer frequency of production (which almost doubled again). The *let* construction is by far the most frequent one, while the rarest patterns include demonstrative pronouns in causee

positions and specific infinitival patterns with less commonly considered periphrastic causative verbs.

- The key observation here is that the distribution of different periphrastic causatives shows a strong similarity across all three age groups studied. The patterns that are most common in adult speech are also the most common in early speech, both in the 0-3 and in the 4-6 age groups. Similarly, patterns with the lowest input frequency are not demonstrated by children in the younger age groups, indicating the absence of a particular need to produce patterns otherwise underrepresented in the input. Also, the production volume of the pattern *help*-ProPers-V was high in the 0-3 age group, closely approaching that of other groups, which can be anticipated in the 4-6 age group, as their overall production of periphrastic causatives is similar to that of adults, but it is unexpectedly high in the 0-3 age group as well, especially compared to other patterns within that age group.
- The findings regarding the less common infinitival realization of the periphrastic causative construction are particularly intriguing. Despite analysing a wide range of verbs with causative implications (43 in total), only three other different verb lexemes (aside from *help* and *get*) were observed in both the 0-3 and the 4-6 age groups. In child-directed speech, these verbs were dominant in terms of utterances but were not the only ones used. It is evident that the most frequently encountered verbs by children are the ones they heavily rely on in their early speech, thus corroborating the usage-based claims on early developmental trajectories of language.
- There is a significant overlap in the frequency of usage of the main verbs in the periphrastic causative constructions among all three age groups. While the predominant verb used is *let* comprises the majority of their usage, the remaining verbs also hold a similar position in terms of their usage hierarchy. However, it is noteworthy that there is a higher concentration of periphrastic causatives involving the most frequently used verbs in the 0-3 age group compared to the other age groups, suggesting a more prominent item-based trend in their production.
- There is a striking similarity between the three age groups in the most frequent nouns in the position of the causee, even when it comes to items not typically expected in the position (for example, non-animate items). Children's frequent use of said items is most likely impacted by the exposure to corresponding patterns in the input.

- The data on the proportion of lexical items in terms of input/output overlap shows a greater correspondence between the 0-3 age group and adults in comparison to the 4-6 age group and adults. The findings suggest a more conservative use of language in the 0-3 age group, which also agrees well with the suppositions of the item-based character of early language development. The data confirms the first hypothesis across several levels, both when it comes to the structural and the lexical character, but also when it comes to the more pronounced item-based production in the youngest age group.
- Adults use inanimate objects more frequently compared to younger age groups, particularly when compared to the 0-3 age group. Inanimate objects are commonly used across all age groups, but adults exhibit a higher concentration of their usage. The input children receive permits the use of inanimate items in the position of the causee to a surprisingly high degree, which should in turn result in similar numbers when it comes to the early language output. However, proportion-wise, the youngest age group (0-3) more frequently uses inherent animates than inanimates (as opposed to adults), which fits into Tomasello's perspective on imitative learning that is supplemented by various cognitive mechanisms such as distributional learning. In other words, as predicted by the second hypothesis, the character of the slots early on tends to be determined by the most prototypical items in the input, which is reflected in the animacy distribution within age groups for the observed construction slot.
- The measures of lexical diversity across age groups also indicate a mild rise in sentences containing causative *make*-NP-V constructions, while the measures of lexical density stay stable. When all measures are interpreted together, the increase in diversity, regardless of its extent, indicates a rise in different word forms averaged per the number of tokens as the age progresses, which partially confirms the hypothesis predicting an increase in lexical elaborateness despite inter-age group similarities within specific constructions.

5. RELATIVE CONSTRUCTION: CORPUS-BASED COMPARISON OF CHILD AND ADULT SPEECH

The subsequent sections of this chapter delve into the theoretical underpinnings of relative constructions, incorporating a detailed explanation into how the construction has been defined prior to the extraction form the corpora. The theoretical background also encompasses a thorough exploration of prior research, particularly in the context of language acquisition and development (Section 5.1). The chapter then proceeds to introduce the acquired data, presenting results inter- and intra-age group analysis of the construction's lexical and structural intricacies (Section 5.2).

5.1. Operationalization and previous research

5.1.1. Defining relative clauses and sentences

Before proceeding with the operationalization and the definition of the 'relative construction', we first need to explore more familiar and prevailing concepts of 'relative sentences' and 'relative clauses' as all of them are closely interconnected. When it comes to relatives, and English relatives in particular,

Different grammar books adopt different definitions of relative clauses and sentences. One can insist on distinguishing between the two terms, in so much that a relative clause signifies solely a subordinate clause, while the relative construction may refer to the entire utterance (consisting of both main clause and a subordinate one). Mel'čuk (2021) provides a distinction between the terms 'clause', 'sentence' and 'construction': (1) the term 'clause' is used to denote a subordinate clause functioning as a modifier; (2) the term 'sentence' refers to the entire sentence that contains the relative clause, i.e. it includes both main and the subordinate relative clause and (3) the term 'construction' signifies a noun followed by the subordinate clause that modifies it, all of which together constitute (or behave like) a noun phrase. Nonetheless, most of these definitions revolve around the fact that subordinate relative clauses serve to modify something in the main clause or the main clause itself; for example, common definitions of relative clauses in grammatical textbooks share the following sentiments:

> "A relative clause is a group of words that describes a noun or noun phrase. Relative clauses are also referred to as adjective clauses because the function of these clauses is to describe or modify a preceding noun phrase. In other words, relative clauses describe or provide information about someone or something in the main clause, very similar to the modifying function of adjectives. Like adverbial clauses, relative clauses are a type of subordinate clause and cannot stand alone." DeCapua (2008, p. 313)

The definition is then often expanded by the fact that relative clauses cannot stand alone, but require the main clause, as well as that they typically rely on relative pronouns (such as *who, whom, which, that, whose*) to tie the two clauses together and convey the desired relation. Furthermore, the textbooks may introduce a formal interpretation of relative clauses, or formal types, and a relational interpretation, or relational types of relative clauses (Huddleston & Pullum, 2002, p. 1033). The formal terminology used with regards to the relative clauses is based on more immediate observations, i.e. the formal types can be determined according to the use of specific wh-relative words such as *which, who, whom*, the relational subordinator that, or simply the omission of either of these subordinators (ibid., 2002, p. 1037):

(10) (a) *This is the letter [which drew our attention to the problem].* [wh relative]
(b) *This is the letter [that drew our attention to the problem].* [*that* relative]

In the two examples above (10), both sentences take the NP *the letter* as the subject. This is also called 'relativized element', meaning that it is the referential element in the sentence to which the whole relative clause is tied to. In this case, it is the subject and the relative clause serves to modify the subject of the main clause. The important thing to note is the role of the relativizer in the relative clause. Again, in the examples above (10), the relativizer occupies the role of the subject in the subordinate clause (especially the relativizer *which*, which satisfies all semantic requirements to take on the role of the subject). In cases where the relativizer occupies the role of the object in the defining clause (see below), it can be omitted:

(11) (a) *They served us a tea that nobody fancied.*
(b) *They served us a tea (X) nobody fancied.*[67] [bare relative]

[67] In the context of language processing and acquisition, it is worth noting that some studies indicate that users prefer the form that contains the relativizer as opposed to the bare relative clause, because the inclusion of the relativizer alleviates production difficulty (Race & MacDonald, 2003).

One of the first distinctive points about relative clauses in various grammar books is the qualification of their restrictiveness. Relative clauses are often labelled as 'restrictive'/'defining' and 'non-restrictive'/'non-defining'. In some English grammar books, they are also referred to as essential and nonessential clauses (se DeCapua, 2008, p. 314). The usual approach to this distinction is grounded in discussions on whether the clause is essential for the sentence to have complete meaning or not. When the clause constitutes an integral part of the sentence it is called defining, and when the information conveyed is either optional or redundant, the clause is called a non-defining relative clause. Nonetheless, this rather simplistic approach to the distinction has been altered or elaborated with a more linguistically sensitive take on the issue. Here, the term 'restrictive' relative clause refers to cases in which the subordinate clause directly modifies the relative 'head', narrowing the general superset to a particular subset to which the action in the relative clause pertains (taking on the role of the subject or the object of the verb in the relative clause), whereas 'non-restrictive' relative clauses do not engage in restricting the 'head' but simply provide additional information that serves as a general qualification of the noun-phrase they modify. Although the distinction plays less of a role in spoken language, the difference is often exemplified orthographically, that is, the punctuation changes depending on the category of relative restrictiveness. When there is a comma preceding a relative clause, the clause is a non-defining one. On the other hand, no punctuation is necessary when the relative clause is a defining one, since the information is an essential one for understanding the speaker's meaning. The examples from Fabb (1990, p. 57) illustrate the distinction between restrictive and non-restrictive qualification of the relative head, where (12a) represents restrictive clauses and (12b) non-restrictive clauses:

(12) (a) *The swans which are white are in that part of the lake.* (restrictive/defining)

(b) *The swans, which are white, are in that part of the lake.* (non-restrictive/non-defining)

The implication in sentence (12a) is that not all *swans* are white in general, and only those *swans* which are white reside in that particular part of the lake. The category of items signified by the noun phrase in question (*the swans*) become restricted to just a part of that category to which the relative clause pertains. The implication in sentence (12b) is that all swans are white in general, which means that the category of items has not been restricted, but rather elaborated. Similar distinction, where the change of the referent is additionally emphasized, is perhaps even more evident in the juxtaposition of relative clauses modifying a single nominal

head and the entire clause. Regardless of whether we treat the main clause as the head of the non-restrictive relative clause, or if we simply choose to state that the syntactic relation of the clause to its antecedent is absent in non-restrictive relative clauses, there are evident implications on the semantic interpretations between the two alternatives.

(13) (a) *Jason kept on telling jokes, which made all of us really angry.*
(b) *Jason kept on telling jokes which made all of us really angry.*

(adopted from DeCapua, 2008, p. 315)

In the example 13a, the inclusion of a comma suggests that the relative clause actually modifies the entire proposition (*Jason kept on telling jokes*). In the example 13b, the relative clause modifies a single noun (*jokes*) rather than the entire main clause. The difference is that in the example 13b it was the content of the jokes themselves that angered the audience, while in the 13a it was the entire act of telling jokes (i.e. the fact that Jason kept on telling jokes regardless of their content) that ultimately made everybody angry. Fabb (1990) argues that the main distinction is actually in the fact that restrictive clauses are modifiers and non-restrictive clauses essentially are not, as well as that the relative clause in the sentence (13a) actually does not have a syntactic relation to its antecedent. According to Fabb (1990), the relationship between the head of the restrictive relative clause and the relative clause itself is one of predication, where the modified noun and the relative clause represent co-indexed sister nodes as projections of the same noun phrase; this scheme is rejected for the non-restrictive relative clauses. In other words, while the restrictive clauses are easily discerned as modifiers within the noun-phrase that encapsulates the nominal and the clause, it is difficult to find a hierarchical location in the syntactic tree for the non-restrictive relative clause. Fabb observes that a non-restrictive relative clause is not c-commanded by anything within the hypothetical noun phrase which encapsulates the antecedent (the bare head) and the clause, nor by the noun phrase itself (ibid., pp. 69-70).

Mel'čuk (2021) disagrees with purely syntactic definitions of relative clauses which tend to focus on their internal structure (one that typically includes a wh-relativizer) and which disregard their external syntactic and semantic functions in the sentence. As opposed to Fabb (1990) and others, he suggests that the proper way to understand relative clauses is to observe them primarily (and solely) as modifiers, i.e. we should call relative clauses simply 'modifier clauses' since their essential function is to provide additional information on the relativized

'head'. A complete definition of relative clauses should thus encapsulate their semantic function which belongs to a superset of modifiers, and its syntactic function, which is specified by the form particular clausal type that carries out its purpose (Mel'čuk, 2021, pp. 239-240).

Depending on the dynamics of the relationship between the relative construction and the overall sentence, Huddleston and Pullum (2002, p. 1034) suggest the division of relatives into four types: (14a) integrated, (14b) supplementary, (14c) cleft and (14d) fused. They used the following examples for illustrating these relational types:

(14) (a) *The boys who defaced the statue were expelled.* (integrated relative)
(b) *My father, who retired last year, now lives in Florida.* (supplementary relative)
(c) *It was Kim who wanted Pat as treasurer.* (cleft relative)
(d) *What you say is quite right.* (fused relative)

The first two examples correspond to the distinction between restrictive and non-restrictive relative clauses. The label integrated relatives emphasizes the integrated nature of the subordinate clause in terms of prosody and the necessity of the conveyed information. Supplementary relative clause, like the non-defining one, adds extra information to its antecedent, but besides information redundancy, it is emphasized that the relative clause is not completely integrated into the structure of the main clause. Although seemingly just as complex, the last two relative types express only a single proposition. The cleft relative construction emerged from the more elementary construction (*Kim wanted Pat as a treasurer*), dividing it into two clauses of the so-called "it-cleft" construction. In Diessel and Tomasello (2000, p. 136), the cleft relative is also referred to as the presentational copular clause, where the elements introducing the relativised noun phrase are almost redundant as they do not contribute to the meaning of the overall sentence.

The last example is particularly problematic given that the relative clause cannot be separated from its antecedent like in the other examples. The fused relative basically lacks the relativised element and the entire relative phrase basically functions as a complement in the main clause. In other words, this particular relative type is not always clearly (or consistently) included amongst relatives in general. This relative-like construction has been labelled differently depending on the textbook, whether as complement clauses (Gast & Diessel, 2012), nominal relative clauses (Chalker & Weiner, 2003) or fused relatives (Huddleston & Pullum, 2002):

(15) (a) *I don't think [that he will ever change].*

(b) *[What you say] is quite right.*

The issue here is actually related to the insistence a relative clause has the syntactic requirement to modify a particular nominal (Sag, 1997, p. 444). In the aforementioned cases, there are no clear entities to which the subordinate clause is tied to, and the relations are not obviously comparable with canonical relative clauses. The examples even clearly differ between themselves, as the subordinate clause in (15a) functions as a verbal projection in the form of adverbial, whereas in (15b) the entire subordinate clause serves to replace a noun phrase in the position of a subject. The issue with defining relative clauses is not completely resolved even when they are restricted to nominal modification. The question becomes whether the nominal that the relative clause modifies has to be overt or explicit, i.e. can we regard a clause a relative one if the relativized element was simply covert or implicit, if it was omitted in for the purposes of easing the speech production process (e.g. *I am aware [that they hid the homework]* vs. *I am aware of <u>the fact</u> [that they hid the homework]*). Moreover, since English grammar can treat clauses (especially subordinate ones) as nominals in either biclausal or multiclausal constructions (or as nominal replacements), perhaps there is no reason to maintain the argument that the relative clause necessarily has to relativize a noun phrase, and not the clause (e.g. *They used to write their homework diligently, which was fortunate for us*). In a more flexible definition of relative clauses, one must only look for a relativized element (regardless of its immediate form) in a biclausal construction to find a relative sentence. The fact that an adverbial clause tends to "take the form of, or share properties with, relative clauses" is noted by Thompson, Longacre and Hwang (1985). They give examples of adverbial clauses indicating time, location and manner and paraphrase them in a way that they syntactically correspond to relative clauses in English, paraphrasing sentences such as *We'll go when Tom gets here* (time), *I'll meet you where the statue used to be* (locative) and *She spoke as he had taught her to* (manner) into *We'll go at the time at which Tom gets here*, *I'll meet you at the place at which the statue used to be*, *She spoke in the way in which he had taught her to*. They further explain:

> "[…] we can see that the relative pronoun referring to the place functions as the location in the relative clause, and the noun phrase the place at which the statue used to be functions as the location in the main clause. In other words, time, locative, and manner clauses state that the relationship between the time, place, or manner of the event in the

main clause and that of the subordinate clause is the same. And it is precisely for this reason that they often share properties with relative clause constructions." (Thompson, Longacre & Hwang, 1985, pp. 178-179)

Nevertheless, in order to avoid confusion, this thesis restricts itself to the traditional approach to relative clauses and constructions, which imply the existence of an overt noun phrase to which the relative subordinate clause attaches itself. In this context, Gast and Diessel (2012) provide a set of characteristics based on which we can more easily determine which label corresponds best to different types of subordinate clauses. The most important distinctions here are that between complementation, coordination and adjunction, as well as those between verbal and nominal projections. According to Gast and Diessel (2012), the examples in (6) function as complement clauses rather than adjunctions, while the relative clause can only be considered as such when it constitutes an adjunction. Complement clauses fill the position of an argument of a predicate, thus occupying the essential syntactic position indispensable for completing the meaning of the sentence. In contrast to complement clauses, adjunct clauses attach themselves to the antecedent such as noun phrases or entire clauses but without being syntactically essential for the entire construction (although they may be semantically and pragmatically essential as in the defining relative clauses). In this context, we can distinguish between 4 major types of subordinate clauses. These are: relative clauses, nominal complement clauses, adverbial clauses and verbal complement clauses. Two of these types (relative clauses and adverbial clauses) function as adjuncts, and two function as complements (nominal complement clause and verbal complement clause) (Gast & Diessel, 2012, p. 6):

(16) (a) Relative clause (adjunction: nominal projection): *the house [that you bought]*
(b) Adverbial clause (adjunction: verbal projection): *He's angry [because she left]*
(c) Nominal complement clause (nominal projection): *the fact [that he was angry]*
(d) Verbal complement clause (verbal projection): *He said [that he was angry]*

The distinction that is problematic for this research is the distinction between the nominal complement clause and the defining relative clause (which would, according to these subtypes, belong to relative adjunct clauses). The only difference between the examples stated may be found in the semantically empty antecedent *the fact* (in the nominal complement clause) and *the house* (relative clause), which is a referent with a clear meaning and signification. In this case, the distinction is also a matter of constructionalization, whereby the noun *fact*

constitutes a part of a construction that is structurally relative, but semantically a complement clause (cf. Traugott & Trousdale, 2014). Since there may be some grey area between the two, it is necessary to adopt a more flexible definition of a relative construction in the context of language acquisition, and the corpus-based exploration used in this research. In other words, this research includes both examples of nominal projections as relative clauses, partly because the extraction of instances from the corpora can hardly be that sensitive on a large-scale dataset, and partly because it is questionable whether these subtle differences play any role in the children's early acquisition of more complex syntactic constructions.

In the context of language acquisition studies, the most important distinction has been between subject and object relative clauses, and many studies have focused on the differences between the two in terms of comprehension and production difficulties (Gennari & MacDonald, 2008; Race & MacDonald, 2003; Diessel & Tomasello, 2005). The difference between subject and object relative clauses can be seen in the following examples (from Gennari & MacDonald, 2008, p. 162):

(17) (a) Subject relative: *The reporter that attacked the senator admitted the error.*
(b) Object relative: *The reporter that the senator attacked admitted the error.*

The difference between subject relative clauses and object relative clauses is in the role that 'the reporter' serves in the subordinate clause – it serves as a subject in the first sentence (17a) and as an object in sentence (17b). The differences in the syntactic arrangement and roles also turned out to play major roles in terms of processing and understanding (see Section 5.1.3 for previous research on the matter).

Other approaches to relative clauses went further and distinguished between 4 types of relative clauses depending on the evident syntactic analysis of the clauses in question, i.e. they differ in their relation towards objects or subjects in the main clause of the sentence. This approach opens up the possibility of discerning between the following 4 types of relative constructions: (a) relative clauses that modify the main-clause subject and the relativizer takes its role (SS relatives), (b) relative clauses that modify the main-clause subject where relativizer becomes the object in the subordinate clause (SO relatives), (c) relative clauses that modify the main-clause object where the relativizer becomes the subject in the subordinate clause (OS relatives), and (d) relative clauses that modify the main-clause object where the relativizer takes

its role as the object again in the subordinate clause (OO relatives) (Diessel & Tomasello, 2005, p. 882). The following examples are taken from Tavakolian 1977:

(18) (a) *The horse [that pushed the goat] stands on the lion.* (SS)
(b) *The cow [that the sheep pushed] stands on the kangaroo.* (SO)
(c) *The cow pushes the kangaroo [that jumped over the goat].* (OS)
(d) *The kangaroo stands on the pig [that the sheep pushed].* (OO)

Naturally, the types further expand with the relativizers such as *whose* into relative clauses where a genitive attribute is gapped or relativized (e.g. *[…] the lady whose daughter goes to the church*).

Apart from the traditional relativizers (*that, which, who, whom, whose*), some textbooks also add question words (labelled as relative adverbs) as relativizers in certain syntactic environments (De Capua, 2008, p. 327-330). However, the addition of these creates problems as it invites the discussion on the differences between adverbial and relative clauses.

5.1.2. Defining the relative construction

When compared to ditransitive and periphrastic causative patterns, it seems less common to see relative patterns being referred to as 'constructions'. Depending on what they refer to – complete biclausal (or multiclausal) sentences or subbordinate clauses – they tend to be addressed as either relative sentences or relative clauses much more frequently. In the same way that other clauses are seldom referred to as construction (e.g. interrogative or declarative clauses), relative clauses arguably constitute a more comprehensive grammatical category than constructions which are characterized by more stable syntax and a meaning attributed to it. Nevertheless, the relative patterns do exhibit some important restrictions in terms of both syntax and semantics, as well as externalize a rather homogenous and unvarying function in their use. Although Culicover (2011) challenged the use of the term 'construction' for English relatives, he did maintain that conceptually they constitute a proper way to describe grammatical phenomena:

> "The main point, which echoes the perspective of many if not all constructionalist approaches to grammar, is that constructions do not differ from one another in

completely arbitrary ways, but cluster around the regular correspondences of a language, while maintaining their special distinguishing properties of form and function." (Culicover, 2011, p. 1)

There has been some inconsistency in denoting English relatives and, even in *The Cambridge grammar of the English language*, we find the authors briefly discussing the overlap between the terms 'relative clause' and the 'relative construction', where the latter term represents a broader category that encompasses cases with the implicit head noun in the relativizer *what*, otherwise not considered necessarily as relative clauses (see Huddleston & Pullum, 2002, p. 1033).

According to other accounts, the relative clause is defined in relation to the relative construction. For example, the relative clause can be described as a subordinate clause that accommodates a variable bound by the 'head' of the relative construction (also referred to as the 'relative head' or the 'head noun'), where the head typically represents an antecedent in the form of a noun phrase (De Vries, 2018). This noun phrase is the one that relativized by the relative clause, which can have restrictive or non-restrictive properties, but it is always bound to the head of the relative construction. Furthermore, the term 'relative construction' is sometimes extended to cases where the head is tacit, that is, where the relativizer is considered to implicitly contain the 'head noun'. This implicit antecedent is common to the relativizer *what* and although the attached clause corresponds to the adverbial or the interrogative one, there have been some arguments about the distinction between the two. An example of a headless or free relative construction would be *what Anne had bought* in *Peter admired what Anne had bought*, as opposed to the same but now interpreted as an embedded question in *Peter asked what Anne had bought* (De Vries, 2018, p. 2).

The distinction that has been typically used to discern relative clauses according to their modifying role in terms of information relevance has also been expanded to denote relative constructions, precisely because the term construction encapsulates the 'head' of the relative clause. Mel'čuk (2021) has also used the terms 'restrictive' and 'descriptive' for 'relative clause constructions', and the definition is basically the same as for 'defining' and 'non-defining' (or 'restrictive' and 'non-restrictive' relative clauses). Restrictive relative clause construction enriches the meaning of the head in the sense that it narrows the referential scope to the item in question by specifying precisely what distinguishes the relativized item from others.

Descriptive relative clause constructions provide additional information on the referent, but the information does not directly impact the reference of the head (or at least the perception of its role in the wider context).

The objections to propositions of something like 'relative clause constructions', or 'relative constructions' in general, come from Culicover (2011). In his work, he emphasizes the individual differences across different relative constructions depending on the choice of a particular relativizer. He states several different morphosyntactic properties make the matter rather incongruous: for example, relatives can be finite, infinitival or participial; they can relativize both NPs and non-NPs; not every relative clause has overt referents and not all relatives can be regarded as clauses at all. His overall claim is that the term 'construction' needlessly overcomplicates the issue, and that not every linguistic phenomenon has to be described in terms of universal syntax and semantics ascribed to it. Perhaps there may be nothing inherent to the relative construction per se, which means that the term 'clause' remains perfectly suitable. In other words, "relative clauses are clauses, and inherit certain properties in virtue of this" (ibid., p. 12), implying that there are differences between English expressions with a particular and stable set of characteristics and the term 'construction'. Consequently, in order to label something as a 'construction', one needs to distinguish generalisations deduced from universals that do not have to be specified for each particular construction and those idiosyncratic properties which arise only as a consequence of some particular (and narrow) phrasal configuration.

An important work in this regard was put forward by Ivan A. Sag (1997) who proposed the term 'English relative clause constructions', and who outlined a number of syntactic constraints exclusive to relative clauses in comparison to others, as well as the fraternal interrogative clauses (e.g. the interrogative and perfectly legitimate *I wonder [who(m) to place my trust in]* (interrogative) vs the inadmissible relative **the baker [who(m) to place your trust in]*). The three fundamental constraints are inherent to the relative construction: the first constraint is that relative clauses cannot stand independently unlike other clauses (such as declarative, interrogative or imperative ones); the second constraint is that they never manifest an inverted word order, and the third constraint is that they must modify a noun (Sag, 1997, p. 444). He defended the idea that relative constructions are indeed to be treated as types, which was the interpretation he deemed relevant in the context of language acquisition study:

> "The results obtained here are of further interest with respect to issues surrounding the innateness and task-specific nature of human knowledge […] If it is indeed possible that even phenomena as grammatically complex as relative clauses can be analysed in terms of a type system like the one presented here, then perhaps less of language has to be thought of as 'hard-wired'. That is, if the program outlined here can be sustained, namely that linguistic knowledge consists of just a system of types and associated constraints, then perhaps much of the nature of grammars can be explained in terms of general cognitive principles, rather than idiosyncratic assumptions about the nature of the human language faculty." (Sag, 1997: p. 478)

In order to remain consistent throughout this work, the relative pattern will henceforth be referred to as relative construction, but it needs to be distinguished from the terms 'relative sentence' and 'relative clause'. In some sense, the term represents a borderline case between relative sentence and relative clause, as it refers to the nominal element that is immediately followed by a relativizer and the rest of the relative clause (see Lehmann, 1986; Schmied, 1993; De Vries, 2018, Okugiri, 2014; Grosu, 2012 etc.). In this thesis, I adopt the definition provided by Lehmann (1986, p. 664):

> "A relative construction is a construction consisting of a nominal (or a common noun phrase, in the terms of categorial grammar) (which may be empty) and a subordinate clause—interpreted as attributively modifying the nominal. The nominal is called the head and the subordinate clause the RC. The attributive relation between head and RC is such that the head is involved in what is stated in the clause."

In other words, the relative constructions targeted by this research cover all nominal projections with an overt head and an overt relativizer, including those instances where the semantically empty noun phrases occupy the relativized slot (e.g. *the fact that he was angry*). It also takes into account cases with relative adverbs *where* and *when* as relativizers which anchor themselves to nominal antecedents (e.g. *I loved the times when we went skiing*), including borderline instances where pronouns are involved (e.g. *I loved it when we went skiing*).[68]

[68] Again, there are valid objections to treating constructions with semantically empty heads as relative ones, and the same goes for instances where the existence of the nominal relativized element is more a consequence of the phrasal nature of an expression than the syntactic need for an overt head (cf. Gast & Diessel, 2012). Though *I loved it when we went skiing* arguably qualifies for an adverbial clause because the antecedent *it* can be omitted,

5.1.3. Previous research

The research shows that children begin to use relative clauses at around the age of two (Clark, 2009, p. 234). In order to acquire the relative clause, it is necessary to identify both form and function (identify the referent that the relative clause modifies), and finally to use this construction whilst demonstrating flexibility in the modification of different referents in different syntactic positions. However, it is not always easy to determine whether the initial attempts at producing relative constructions indeed represent 'relative constructions'. According to data, they are frequently incomplete, and in the initial stages of production these constructions lack essential elements that form proper relative constructions such as wh-relativizers. In cases of initial production where the construction lacks certain elements, researchers analyse them carefully and decide whether or not to treat them as relative construction. For instance, they may choose to ignore the fact that the sentence lacks the necessary relativizer and hypothesize that the child did in fact have the complete relative construction in mind but could not produce it due to other reasons unrelated to comprehension of linguistic structures. This is corroborated by examples of sentences produced one month after reaching the age of two; the utterances are said to reveal the incremental process behind the acquisition of relative constructions (*That map gorilla live*.; *Herb work big building have elevator'n it.*; *That birdhouse bird lives*.) (Clark, 2009, p. 234). However, classifying incomplete constructions as relative constructions poses certain theoretical problems for linguists working within both cognitive and nativist paradigms. The evidence that these sentences are indeed representations of relative constructions, regardless of being incomplete, is found in the fact that they are followed by the production of full relative constructions only several months after the initial efforts.

The problems associated with early language production and distinguishing relatives from other similar linguistic patterns have only served as a stepping stone to another important discussion in language acquisition studies – one that concerns syntactic complexity and the way in which the investigators should approach more complex constructions as opposed to

the inclusion of these cases is due in part to the difficulty of accurately extracting instances from large-scale datasets, as well as to the question of whether these subtle distinctions have any impact on children's early acquisition of more complex syntactic structures.

"simpler" ones. One such example comes from Lust et al. (2009), who have argued that there is no difference between the acquisition of relative constructions and other – typically interpreted as more simple – constructions. At the core of their argument is precisely the fact that it is hard to draw a clear distinction between simple and complex sentences in grammatical terms and, consequentially, there should not be any fundamental distinction in the context of their acquisition. In their overview of the research conducted on relative sentences in view of language development and acquisition, the authors claim that the data do not support the position where complex sentences evolve from simpler ones in additive manner. Their position is that evidence points towards the existence of complex sentence grammar from the very beginning instead of relativization emerging as an expansion of simpler constructions.

Nevertheless, the majority of research in relative clause acquisition has oriented towards different types of relatives in relation to subject and object positions within the main and subordinate clause, Thus far, the data available shows that relative clauses are more easily understood by children when the subject of the main clause in the relative construction is also the subject of the subordinate relative clause (e.g. *The songs that I taught him are nice to listen to*) (Clark, 2003, p. 235). The same argument is put forward for relative clauses that directly attach themselves to the object of the main clause but remain subjects in the subordinate clause (e.g. *I taught him to play songs that are nice to listen to*), but it appears more difficult to acquire constructions where the object of the main clause stays the object of the subordinate clause (e.g. *I taught him to play songs that everybody wants to listen to*), or where the relative construction is also a passive one (e.g. *I taught him to play songs that are loved by everybody*) (Waters & Caplan, 1996; Gordon, Hendrick, & Johnson, 2004; Gennari & MacDonald, 2008). The initial hypotheses pertaining to the processing of relative clauses were rooted in the underlying observation that a sentence in which some of its linguistic units have been interrupted or rearranged will be harder to process than the sentence where this does not seem to be the case (Slobin, 1971, p. 352). Following from this, the clauses which modify the subject NPs of the matrix sentence[69] (SS relatives) should be more difficult to process than the ones which modify object NPs (OS relatives), because the matrix sentence is not interrupted in the latter case (e.g. *The man who hugged the woman was John* vs. *They found the man who hugged the woman*). On the other hand, the sentences where the relativized element takes on the role of the object

[69] A matrix sentence is the one which contains an embedded clause as its constituent (A first dictionary of linguistics and phonetics, 1980).

in the relative clause should be more difficult to process because the order becomes rearranged (e.g. *The man that the woman hugged was John*). So far, the research has more or less been in agreement on this matter – indeed, object relative clauses are more difficult than subject relative clauses, at least comprehension-wise (Gordon, Hendrick, & Johnson, 2001, 2004; Just & Carpenter, 1992; MacWhinney & Pleh, 1988; Warren & Gibson, 2002 etc.). In a number of influential studies on children and adult processing of relative clauses that found subject relative clauses to be easier to comprehend than object relative clauses, the reasons have typically been attributed to object relatives' greater syntactic complexity or cognitive requirements (see Gibson, 1998; Miyamoto & Nakamura, 2003; Friedmann & Novogrodsky, 2004).[70] However, this explanation was not always deemed sufficient and other factors pertaining to their more frequent occurrence in the input and greater predictability in natural language started to be examined. Subsequent research has shown that the difficulty of object relative clauses depends on a range of factors, including the specific properties of the language being learned, the object relatives children often hear and say, the length and complexity of the clause, and the cognitive and linguistic abilities of the child (cf. Pickering & Traxler, 1998; Kidd et al., 2007; Tsoi et al., 2019).

In her research that dates back to 1974, Sheldon looked into children's comprehension of English relative clauses in relation to how they process subject and object relative clauses. Her results went against the initial expectations that children would more easily process non-interrupted and non-rearranged relative sentences (see Slobin, 1971), but instead suggested that the crucial element processing-wise was the transposition of the subject NP of the matrix sentence into the object NP of the relative clause (and the other way around); that is, the children were better at processing sentences where the role of the NP was maintained in both matrix sentence and the relative clause (such as SS and OO relative sentences). In this context, it is possible to separate two hypothesis concerning the processing difficulty of relative sentences as follows: (1) OS > SS,OO > SO and (2) SS, OO > OS > SO, where '>' means "easier to parse" or "less complex". In her subsequent study (1977), she compared the data gathered on children with the data newly obtained on adults (students at the University of Minnesota). The data collected on adults showed different patterns in terms of most frequent errors depending on the

[70] Some research looked into the interplay of different factors; for example, Race & MacDonald (2003) investigated whether the use of the relativizer *that* can facilitate the production of object relative clauses, the assumption being that it gives speakers more time to plan the embedded clause, which was ultimately confirmed for oral language production as well.

relative sentence type; the results on the adult sample were in agreement with Slobin's predictions (1971) that the easiest relative pattern to process should be the one which is neither interrupted nor rearranged, or in other words, the OS relative sentence (e.g. *They found the man who hugged the woman*). Regardless of the discrepancy between the two samples, Sheldon (1977) concluded that both children and adults follow the same processing strategies when it comes to these sentences. She claimed that the difference between adults and children when it comes to making errors across the four types of English relatives resides not in different processing strategies, but in the degree to which they rely on these strategies (whether they overuse some as opposed to others). Although both age groups were argued to share the same strategies, it was obvious that children differed from adults in two aspects: (1) the way in which they found the noun phrase modified by the relative clause (children parsed object modifiers as if they were subject modifiers) and (2) the way in which they located the function of the relativized NP (the grammatical function of the relativized NP tended to be interpreted the same as that of its antecedent—thus predicting that would be easier to process sentences when the coreferential NPs have the same grammatical function in their respective clauses).

When it comes to psycholinguistic approaches to relative sentences, a lot of studies had taken particular interest into object relative clauses. A part of the reason for this is that the pattern is considered to require greater processing effort on behalf of both those who produce it, as well as those who hear it. Other relative constructions do not require the initial noun phrase to be retained in the memory as long as this, and then to be readdressed as the sentence unfolds. Some studies have addressed the exact cause of why object relative constructions tend to encounter more comprehension difficulty than the subject relative ones (e.g. see Staub et al., 2017). They found that the underlying cause for a sentence such as *The reporter that attacked the senator admitted the error* to be more easily processed than the *The reporter that the senator attacked admitted the error* resides in the role switching required by the matrix verb that follows the relative construction, as well as the interference between the matrix verb and the relative clause verb.[71] For example, 'the reporter' in the latter construction (object relative construction) is initially identified as the subject of the relative clause, but as the sentence unfolds it

[71] The debate on the underlying causes for greater processing difficulty of object relative clauses is represented through the opposition between thematic bottleneck accounts and retrieval interference accounts, of which the former insists on the role of the matrix verb switching (see also MacWhinney & Pleh, 1988), while the latter emphasizes the interference imposed on the retrieval of the matrix subject by the object relative clause subject (see Van Dyke & Lewis, 2003; Van Dyke, 2007).

transposes to an object due to the requirements of the relative clause verb ‘attacked’; in the end, the comprehender is once again required to reintroduce ‘the reporter’ as the subject of the main clause due to the requirements of the matrix verb ‘admitted’.

Some researchers delved into the cognitive background of relative sentence processing (see Finney et al., 2014; Rusli & Montgomery, 2017). A study by Finney et al. (2014) looked into the role of working memory and attention focus switching in the processing and comprehension of object relative clauses[72] in 7-11 year old children. Working memory storage was predicted as important due to the necessity for children to retain the two unintegrated NPs in memory during sentence processing, and the ability to switch attention was seen as potentially relevant because the comprehension of object relatives requires one to reactivate the initial noun phrase, and thus momentarily switch attention from sentence processing towards memory retrieval. In the end, their results suggested that both working memory and attention focus switching were important for children’s comprehension of the object relative clauses. Rusly and Montgomery (2017) also looked into the role of working memory in the processing of object relatives on a similarly aged sample (9-11 years of age), but contrasted it with the role of language knowledge (their vocabulary prowess). Their findings did not suggest that working memory plays a role in object relative processing, but children’s lexical familiarization with the language explained for more than 20% of the variance in their processing of object relatives.

On the other hand, in more linguistically-oriented research, the processing differences were examined with respect to structural and lexical features of relative constructions. One of the earliest attempts was to attribute some of these differences to animacy. For instance, Goodluck and Tavakolian (1982) conducted experiments on 4-5 year old children and found that animacy plays a role in children’s comprehension of OS relative sentences, that is, the object’s animacy in the relative clause (e.g. *the sheep* in *The dog kicks the horse that knocks over the sheep*) affected whether children would understand the roles of the participants in the sentence. The researchers collected data based on their task where the children were asked to act out the sentences presented to them with the help of toys, and after conducting the experiment, they concluded that a higher proportion of errors was recorded with sentences that contained animate objects in the final position (e.g. animate *the horse that knocks over the sheep*

[72] They tested for sentences such as *Bob fed the camel to which the pink lobster showed his new computer game at his office on Monday morning* (2014, p. 3).

vs. inanimate *the horse that knocks over the table*). They assumed that the reason for this might be due to the increase of the processing load that the animacy of all participants incurs on the comprehender. Moreover, they found a considerable difference between 4 year olds and 5 year olds in the number of errors, with older children providing 67.5% and four-year-olds 59% correct responses to the task at hand.

In their study aimed at investigating how children produce restrictive relative clauses, McKee et al. (1998) examined 28 children between the ages of 2 years and 2 months and 3 years and 10 months. They focused on restrictive relative clauses, which they deemed as structurally complex and as such, more important in the study of child language development. The majority of the children produced mostly adult-like relative clauses during the experiment. However, when they did makes mistakes, these errors typically involved the head (non-matching context or the missing head), number (agreement problems), auxiliaries (e.g. omission or addition), or pronominal elements (e.g. the use of *what* as a relativizer) of the relative clauses. Despite these errors, the study found that children were able to represent the syntactic structure of the embedded clause accurately at an early age. After discussing the implications of the results, McKee et al. (1998) highlighted the importance of examining production data in understanding how children acquire language. The study by Kidd and Bavin also examined how 3- to 5-year-old English-speaking children understand sentences with restrictive relative clauses, and found that children acquire right-branching structures before center-embedded ones. The study also suggests that there are both general cognitive and language-specific constraints on language development, and English-speaking children experience more difficulty processing center-embedded structures compared to right-branching structures. The avoidance of center-embedding structures in English may help the language processor parse sentences more efficiently, as the main clause in a right-branching structure occurs before the relative clause and can be processed independently. According to them, this would explain why children in the study performed equally well on the OO and OS constructions.

Another study that looked into differences between English monolingual children and adults in terms of processing difference in relative sentences comes from Felser et al. (2003). They compared 6-7 year old children to adults in their processing of ambiguous relative sentences with two potential attachment sites for the relative clause (e.g. *Someone shot the servant of the actress who was on the balcony*, where it is not clear whether the relative clause modifies the *servant* or the *actress*). In their comprehension of the ambiguous relative reference,

adults tended to rely on the semantic properties of the preposition joining the relativized NP (the examined relatives modified NPs joined by either *of* or *with*), while the children attached the relative clause to either first or the second antecedent depending on their listening span ability; that is, children with greater listening spans preferred to attach the relative clause to the initial antecedent (head of the relativized noun phrase), whereas those with shorter spans tended to attach it to second noun within the phrase. The authors of the study also argued against the so-called Tuning Hypothesis (cf. Mitchell and Cuetos, 1991; Cuetos et al., 1996), whereby the input plays a crucial role in the comprehenders' preferences with respect to NP attachment in relative clauses, claiming that the children would have behaved in a relatively uniform way had this been the case (if one assumes a balanced input with similar proportion of NP1 and NP2 attachments), rather than being influenced by their listening spans in their attachment preferences – as the data seem to have indicated.

Due to complexity and various alternatives in terms of relative sentence structure, even in the context of development, the research has not been solely restricted to subject and object relatives dichotomy, branching and restrictiveness. For example, there have been studies that have evaluated the treatment of resumptive pronouns[73] in the context of English relative clause acquisition (Pérez-Leroux, 1995; McKee & McDaniel, 2001). McKee and McDaniel (2001) looked into differences between children (aged 3;5-8;11) and adults; they found similarities between children and adults in view of their production data, while their judgment data indicated that children tend to overaccept resumptives in sentence positions where the resumptive pronoun can be extracted without hurting the grammaticality of the sentence. However, the authors did not interpret this discrepancy between children and adults in a framework that suggests children's grammars to be immature when compared to those of adults, but rather claimed that these differences emerged as a by-product of specific sentence processing, and that children demonstrate adult-like grammar with respect to resumptive pronouns. This line of reasoning was different from that of Pérez-Leroux (1995) who, in the same context of English resumptive pronouns, found children's grammar to be considerably different from the one of adults.

[73] Resumptives designate pronouns used to reintroduce items which have already been realized within the same expression, but they tend to occur marginally in the English language, such as in English relative clauses (e.g. *I had some other point which I can't remember what it is*; cf. Chomsky, 1982; Cann et al., 2005).

In the overview of his previous work with Michael Tomasello, Diessel (2009) summarised the key features with respect to the acquisition of relative clauses.[74] His first observation was that children's earliest relatives tend to be simple in so far as they express a single event or point. Though embedded and forming a complex sentence from a syntactic point of view, the fact is that the proposition in the main clause is typically empty or low-level in terms of transitivity. Furthermore, Diessel stated that the acquisition of the internal properties of relative clauses is influenced by various factors, only one of which (though the most important one) is the frequency of different types of relative clauses in the language children are exposed to (ibid.). For instance, subject and object relatives are very common in English, while genitive (e.g. *The man whose dog bit the woman*) and indirect object relatives (e.g. *The man who the woman gave the book to*) are extremely rare. As a result, children have fewer difficulties with subject and object relatives than with other types of relative clauses. However, input frequency alone is not enough to explain the data, as there is another factor that plays a key role in this study, namely the similarity between different types of constructions. For example, subject relatives may cause fewer problems because they are similar to simple sentences, which children learn before they start producing relative clauses. Object and adverbial relatives[75] could potentially cause the same amount of difficulty due to similar word order as far as the sequence of the relativizer, relative clause subject and verb is concerned (e.g. *the house [that you bought]* vs. *the house [which you lived in]*). The reason why similarity is important is that relative clauses are grammatical constructions, which means that they are form-function pairings that are related to each other in an associative network like lexical expressions. Children acquire this network in a step-by-step manner by relating new relative clause constructions to those they already know. The process begins with subject relatives, which are not much different from simple sentences and have the same word order as simple main clauses, and it ends with genitive relatives, which are the most distinct from all other types of relative clauses. Finally, Diessel (2009b) referred to the exploration of semantic and pragmatic characteristics of subject and non-subject relatives produced spontaneously by children. The findings suggested that subject and non-subject relatives represent different types

[74] cf. Diessel and Tomasello (2000, 2005); Diessel (2004); Brandt et al. (2008).

[75] Note that the term relative adverbials captures a different relationship from adverbial adjunct clauses defined by Gast and Diessel (2012, p. 6); see Example 16 in this thesis. The term relative adverbial is used to describe constructions where the oblique element such as an adverbial is relativized as opposed to, for example, a direct object of the relative clause (cf. *This is the girl who the boy teased at school yesterday* and *This is the girl who Peter played with in the garden*; see Diessel and Tomasello, 2005; p. 884; also see Diessel, 2009).

of situations. Subject relatives are common after both animate and inanimate nouns (e.g. *the man who knows me* or *the car that drove*), whereas non-subject relatives (object and adverbial relatives) tend to modify an inanimate noun (e.g. *the car that he drove*). Non-subject relatives also typically contain a first or second person pronoun as the subject, which is engaged in a dynamic activity (e.g. *the car that I drove*), whereas subject relatives involve lexical subjects that are often embedded in intransitive clauses or transitive clauses with low transitivity (e.g. *the man that danced*). Diessel finally argued that the differences in the semantic properties of subject and non-subject relatives are indicative of their respective pragmatic functions; subject relatives are usually used to provide more information about a person or thing that was just mentioned in the main sentence (e.g. *the man who loves sandwiches*), while non-subject relatives are mostly used to specify the identity of a person or thing by describing its relationship to a known entity, such as a participant in the conversation or a previously mentioned object (e.g. *the man who I talked to*).

5.2. Relative constructions – results

5.2.1. Extraction method and the examples of relative constructions

When it comes to the extraction of relative constructions, the location of instances in the corpora is centred around specific relativizers used in the relative construction. Similarly to the periphrastic causative, the construction is, at least in part restricted to particular items/relativizers. On the other hand, the POS tagging tool was used to target those instances where a particular noun (or pronoun) is immediately followed by the relativizer. This leaves out some of the instances in English, particularly those where the relativizer ties to the external referent (usually an idea or an event) captured by the entire clause (e.g. *I love spending time on the island, which is uncommon for the rural folks such as myself*). In other words, by mapping the constructions based on POS constraints and constraints pointing to specific lexemes (in this case, different types of relativizers) the relative constructions extracted from the corpus were solely those that relativize a particular noun phrase (see Appendix 10.1 for POS syntax), therefore reaching a higher degree of query precision. Additionally, because the relative pronoun can be omitted in the cases where the relativizer is not the subject of the subordinate clause, the corpus-based search for relative clauses becomes significantly more difficult. The

problem occurs when the search has to be anchored in certain elements, and the only way to ease the search for relative constructions is to anchor them in relative pronouns. In this study, because of the impossibility to additionally improve the precision of the queries and due to the interpretative issues pointed out in Section 5.1.3, the focus was has been shifted to locating those examples of relative construction which contain a relative pronoun, thus leaving the more time consuming analysis of zero-relative clauses to future research.

Some of the examples for the relative constructions found in the 0-3 and 4-6 corpora are listed in Table 27. Even in spoken language, the majority of relative constructions extracted from the corpora (when filtered out) could be considered canonical; it is rather clear what constitutes a relativized noun phrase, what the corresponding relativizer is, and why the relative clause has been employed. Even among the younger children, we find occasional examples of proper relative construction use; for example, two and a half year old children produce sentences such as *Where's the other thing that Grandfather gave me* or *Hey there's a we have some guys who stand up*, where the relative clause is properly used to distinguish the referent from other objects/people designated by the same noun, but generally speaking, there remain more examples which present difficulty interpretation-wise in the youngest age group.

Examples targeted in this research include a wide range of relative construction types with respect to both relativizers and object/subject roles in the sentences. For instance, the examples included cover the constructions that employ the relativizer *where* and *when* such as *Remember this part when they were splashing in the water* or *The inside of it where we sit*, because the relative clauses within are interpreted as elaborations of the relativized noun phrases next to which they stand. In some of the cases, especially when it comes to the locative (where) and temporal relativization (when), there is an overlap between adverbial clauses and relative clauses; in an example such as *Remember this part when they were splashing in the water*, the noun phrase *this part* is syntactically redundant and it is possible to argue that the clause expands on the verb phrase, rather than the noun phrase by answering the question *which part*?. In cases with same (or similar) dilemma, the separating criteria was the overt/covert use of the relativized noun phrase, that is, when expressed and overt, the construction was included in the final dataset. Sometimes, the relative construction was isolated from the rest of the sentence since the contextual character of the spoken language often permits what would otherwise be considered reduced propositions; here, we find examples such as *That doctor who looks after you* or *Fish that he caught already*.

Table 27. Examples of (candidate) relative constructions found in the CHILDES corpora

Corpus	Age (y; m; d)	Produced relatives
DEMETRAS-WORKING	2;2;1	*beep beep [I who wants rocks here?]***
VALIAN	2;5;18	*Where's the other [thing that Grandfather gave me?]*
WEISMER	2;6	*Hey there's a we have [some guys who stand up.]*
VALIAN	2;6;24	*Fish that he caught already?*
WELLS	2;6;5	*There's [a lady in park who's carrying a baby.]*
CRUTTENDEN	2;7;18	*Wait [a minute when I find a domino.]***
SACHS	2;11;12	*I don't like [people who writed.]*
FLETCHER-THREE YEAR OLDS	3;0;24	*He's sleeping on [the flower that one is.]***
LARA	3;3;22	*I mean [one that you had.]*
WELLS	3;6	*That doctor who looks after you.*
KUCZAJ	3;6;3	*The inside of it where we sit.*
WEIST	3;7;4	*Remember this part when they were splashing in the water.*
VANKLEECK	3;9	*That's all the dough that we got.*
HALL	4;6	*Oh those where the chicken were.*
ERVINTRIPP	5;0	*That's a dungeon that they slide down in.*
FLETCHER-FIVE YEAR OLDS	5;0;19	*I think it's a gate who someone broke in to the house.**
MACWHINNEY	5;6;9	*You know this that I was talking about?*
MACWHINNEY	6;9;22	*Yeah Mark that was only at the party.**

*dubious but retained in the analysis **removed from the final dataset

As stated, there are certain constructions which occasionally present difficulties interpretation-wise. But, as with the other constructions, the examples which may be dubious for the interpretation represent a minority of cases, whose inclusion or exclusion does not represent an issue or significantly changes the totality of data. For example, even after filtering out the data, the sentence *Wait a minute when I find a domino* was initially kept as one that contains a relative construction *minute when I find […]*, because one could argue that it answers to the question *Which minute?*. It is only after the second inspection that the sentence has been removed from the final set, because the child probably made a mistake by using the relativizer when as opposed to a more fitting conjunction (such as *while* or *as*). Another example of one such sentence which was subsequently removed is *He's sleeping on the flower that one is*. Initially, it seemed as if *that* relativizes the noun *flower*, but a more detailed examination reveals that the sentence is best interpreted as one with two independent clauses, which should have incorporated a punctuation mark for the pause between what was presumably a relativized noun

phrase and the relativizer. In the example, *that* does not function as a relativizer, but rather as a demonstrative pronoun, with the entire clause functioning as a sort of reiteration of what has been said for the purposes of further emphasis. The case is even less clear with the sentence *beep beep I who wants rocks here?*, for which the surrounding context suggests that it is actually a question *Who wants rocks here?*, and that there should be some punctuation mark for the hiatus between *I* and the rest of the phrase.

Other examples that have been dubious, but were still kept in the final dataset, include sentences such as (1) *I think it's a gate who someone broke in to the house* or (2) *Yeah Mark that was only at the party*. The sentences move away from conventional relatives in so far that they do not give a straightforward impression of how the relative clause and the relativized noun phrase connect in terms of meaning. The first example is a complete miss, as far as the link between the relativizer and the relativized noun phrase is concerned; the child probably intended to say *[...] a gate which someone broke in order to get into the house*, but instead made a mistake of using a relativizer restricted to animates. It is entirely possible that the second example does not represent a relative sentence, but a simple monoclausal statement, following the affirmative phrase 'Yeah Mark!'. Although dubious, these two cases (and few others that were alike) were nonetheless kept in the final dataset, but it is important to note that their impact on the overall data is almost non-existent and hardly corrupts the overall frequencies and calculations.

5.2.2. Structural complexity and similarity

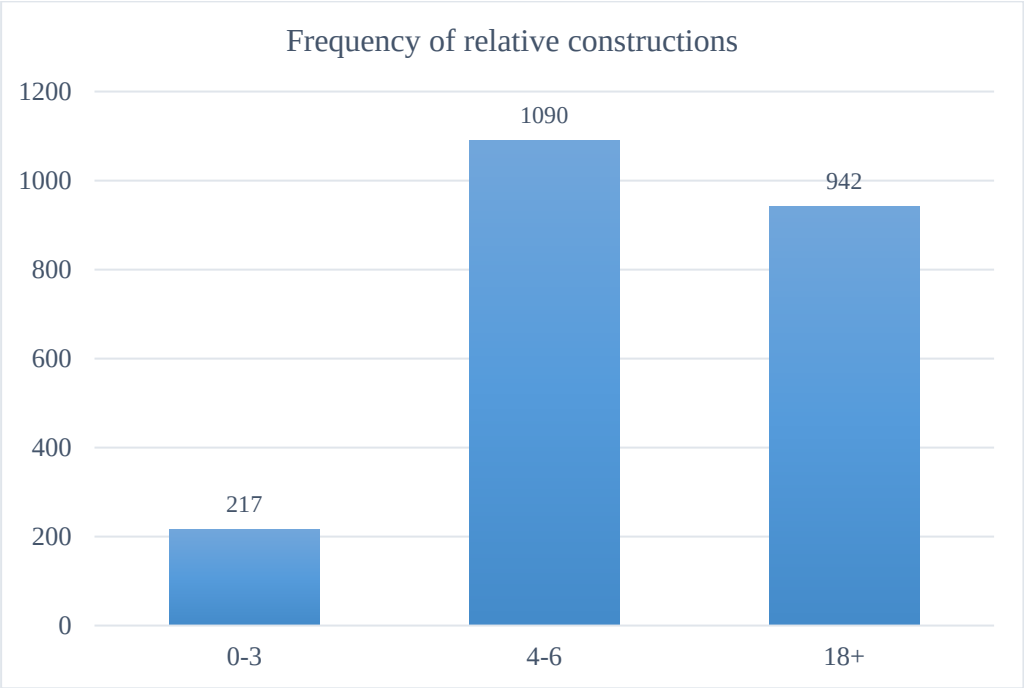
The first look at the data on relative constructions captured by this research reveals (see

Figure 11) that the overall production of relatives is by far the lowest in the 0-3 age group.[76] It is approximately five times lower than in the 4-6 age group, and the case is similar in comparison to the adult group. The most frequent production of relative constructions can be observed in the 4-6 age group, followed by slightly lower numbers in the adult group and considerably lower numbers in the 0-3 age group. To some degree, the overall frequency of production when it comes to relative constructions is surprising, given that the expected rise in

[76] Some of the preliminary results on relative construction sub-pattern frequencies have already been published in Proroković & Angster (2022).

the production of these constructions is actually replaced by a drop between the 4-6 age group and that of adults. As already observed with periphrastic causative constructions, the main reason for this is located in the fact that the adult speech recorded in CHILDES implies a level of linguistic simplification. The fact that a relatively difficult construction tends to be more frequently uttered by the 4-6 age groups than by adults might suggest that the language of adults in these particular corpora represents a simplified version of their otherwise communicatively unburdened speech (in this context, the adults are aware that they are speaking to children, which is why they sometimes might resort to simpler solutions in speech). As already stated, the vast majority of adult utterances in the CHILDES corpus, whether from parents or researchers, constitutes speech directed at children. The research has already shown that this speech is usually significantly "simplified" compared to the usual speech of adults (Snow, 1972; Phillips, 1973; Sachs et al., 1976). For example, adults often use less than four words per sentence as opposed to when they converse with their peers, in which case they use more than eight (Phillips, 1973). The fact that the language aimed at children is simplified does not only refer to the number of words but also to structure; the research by Sachs, Brown and Salerno (1976) has shown that adults use dependent clauses (such as relatives) significantly less, and the same goes for negation. A recent study by Montag (2019) has provided some insight into the character of child-directed speech with regards to the presence of relative constructions in comparison to children's picture books. According to her data, the language in children's picture books was significantly more complex than child-directed language, at least in terms of the total presence of passive and relative constructions which were substantially less frequent in child-directed speech (they were at least five times more frequent in children's picture books, such as with object relative clauses, and the discrepancy was even higher in other types).

Figure 11. Relative frequency of relative constructions across age groups



But what does the frequency have to do with the level of one's linguistic prowess and how may it reflect complexity? These two measures are not always interrelated and they typically need to be treated as a separate phenomena in studies of language use. If someone produces something more often, the reason for this might be linguistically external or simply conditioned by the linguistic medium, as evidenced by countless studies which show noticeable disparities between spoken and written language with a number of different linguistic phenomena (cf. Louwerse et al., 2004; Miller, 2006; Leech et al., 2014; Okugiri, 2014). The point where the two become interrelated (frequency and complexity) is when the researcher treats the complexity of a particular construction in relation to its communicative intent or manner, i.e. in this particular case, it is possible to argue that adults are much more conscious about their language use and the cognitive ability of their interlocutor, and in order to be clear in their communicative intent, they resort to less complex constructions when conveying their message. In the case of relative constructions, this implies that adults indeed perceive relative constructions as a challenge to their interlocutors (for example, more of a challenge than ditransitives) and that they take this fact into account before (or whilst) producing their child-directed utterances.

Regardless, the fact that the use of relative clauses is almost 5 times higher in the age group 4-6 compared to the age group 0-3 is indicative of the children's language developmental

stages, i.e. at least in terms of its key transition periods. This becomes particularly interesting when we take into account the fact that adults (albeit in speech aimed at children) produce less relative clauses than the age group 4-6. The gravity of discordance between the 0-3 and the 4-6 age group is only amplified in relation to smaller difference between the 4-6 and the adult age group, as it seems implausible to suggest that five-year-olds have better grasp of relative constructions than adults. Instead, it seems more prudent to recognize that children start producing relative clauses relatively early, but complete mastery of the relative construction implies children being able to use them whenever the need arises, and with a frequency that corresponds to that of everyday situations and their conversational requirements. The results may indeed be taken as an indication that the acquisitional trajectory slowly reaches its end in the 4-6 age group, at least in view of children not being held back by processing difficulty when producing intended linguistic structures. Therefore, it is certainly necessary to state that children begin to understand and produce relative clauses at an early age, but also that this production is relatively limited. For example, according to Clark (2015, p. 234), children begin to produce relative sentences already around the age of two, although it turns out that these are usually sentences that lack a relativizer (e.g. *I see [ə] building Eve go*.; CLARK, 2015, p. 199). Even the very interpretation of such utterances as relatives proves somewhat problematic and the question becomes whether the key information following the noun *building* arranges as a relative construction in terms of how the expression is represented in the mind of the speaker. In addition, English language speakers characteristically place the additional information in the final position of the sentence, and seldom rely on constructions where the information is inserted within, as the relative constructions in English allow (e.g. *I see a building, which is tall, and where Eve likes to go*). It seems evident from this research that the significant jump in the production of relative clauses comes after the age of four and the evaluation of the extent of this jump may prove useful in comparison to other constructions and their developmental trajectories.

Frequency data according to their syntactic pattern reveals certain relationships in the processing of relative clauses with respect to their grammatical features (Table 28; for the corresponding examples, see Table 29). As far as specific patterns of relative constructions are concerned, once again there seems to be a striking congruence between age groups in how frequently they produce each. Constructions where the bare pronoun takes the relativized position are poorly represented across all age groups, and this especially goes for those with the

relativizer *which* (less than 0.5 instances found per million tokens). Part of the reason for the fact that not many examples can be found for patterns beginning with PRO-*which* may be because *which* is used to relativize inanimate beings, and pronouns more frequently signify the opposite. This would explain the slightly more frequent patterns where the relativizer *who* is adjacent to pronouns, or those with the more flexible *that*. In general, it is clear that the bare pronouns are considerably less frequent than nouns (this issue is explored later; see Table 33), but even among pronoun-based relativisation some exhibit slightly higher degrees of frequency. Both constructions with subject and object relative clauses relativizing pronouns allow the use of the relativizer *that* (see PRO-*that*-NP-VP and PRO-*that*-VP) and seem to occur more regularly than other relativized pronouns.[77] These include utterances such as *This is yours that turn around* (BROWN, 3;5;1), *I found one that used to go in there* (BROWN, 3;8;1), or *I want all of those those that you are putting away* (HALL, 4;6). Another pattern that relativizes pronouns and shows some degree of frequency (at least in the 4-6 and in the adult group) is PRO-*when*-NP-V. This construction does not satisfy the preconditions for the prototypical relative as easily as the rest, as it tends to relativize the pronoun *it* which is loosely interpreted as a replacement for time and is almost always redundant from both a semantic and a syntactic point of view. For instance, these constructions include examples such as *I hate it when you say sister* (MACWHINNEY, 6;7;1), which theoretically can replace *I hate the times when you say sister*, but more likely replaces the adverbial transposition *I hate when you say sister*. The reasons for the rarity of some constructions such as PRO-*where*-NP-VP are quite obvious; in the case of *where*, a pronoun that is relativised by *where* sounds strange when compared to other relativizers.[78] Most of the constructions that happen to be produced least frequently, beside the fact that they relativize pronouns, also tend to be object relative clauses which are more difficult to process than subject relative clauses. Other than this, there is no obvious reason why these would be so

[77] In this context, the label PRO encapsulates both personal and other types of pronouns, such as demonstrative pronouns (e.g. *this, that, these*…) or independent quantifiers acting as indefinite pronouns (e.g. *some, one, all*…).

[78] Consider the following:
I love the place where we played tennis;
I love it where we played tennis.
The latter is not a plausible alternative use of the former, and this constraint seems to be specific to the relativizer *where*, whereas the same alternative works well with the relativizer *when*:
I loved the times when we played tennis;
I loved it when we played tennis.
However, also note that this interpretation of the pronoun *it* as a relativized element is not without its problems, i.e. *it* arguably plays a role of the dummy antecedent in this case, which makes the construction more of a nominal complement clause than a relative one (see Example 16). However, this distinction falls beyond the scope of this thesis, which has explored the production of relative construction irrespective of whether the relativized element is semantically empty or not (see Section 5.1.2 and Footnote 68).

underrepresented in the younger age groups beside the fact that they are seldom found in the input. On the other hand, the most frequent relative patterns (Table 28) naturally converge around nominal relativizations, especially those which employ the relativizer *that*.

Table 28. Distribution of relative constructions across age groups (relative frequencies normalised per million tokens rounded to the closest unit)

Relative constructions	0-3	4-6	18+
PRO-*which*-NP-VP	0	0	0
PRO-*who/whom*-NP-VP	0	0	0
PRO-*which*-VP	0	1	0
PRO-*where*-NP-VP	1	1	0
N-*who/whom*-NP-VP	0	5	2
PRO-*who*-VP	2	2	3
N-*whose*-NP-VP	0	2	3
N-*which*-NP-VP	1	5	5
PRO-*that*-NP-VP	2	7	8
PRO-*when*-NP-VP	0	5	10
PRO-*that*-VP	3	12	14
N-*which*-VP	1	7	23
N-*where*-NP-VP	14	45	47
N-*when*-NP-VP	14	65	76
N-*who*-VP	40	147	129
N-*that*-NP-VP	49	304	280
N-*that*-VP	90	484	341
TOTAL	217	1090	942

Table 29. Examples of the produced relative constructions in the CHILDES corpora (patterns from Table 28 indicated in square brackets)

Construction pattern	Example	Corpus	Age
PRO-*which*-NP-VP	*And I have [football which I don"t like to play].*	WEISMER	5;6
PRO-*who/whom*-NP-VP	*You who have dirt all over your tongue huh?*	BRENT	18+
PRO-*which*-VP	*She wore [this which is a man"s] she wore everything.*	GARVEY	5;2
PRO-*where*-NP-VP	*Ya better not get on [this where the wheels are].*	GARVEY	4;0
N-*who/whom*-NP-VP	*That"s [three people who we have in our room].*	CARTER	6;0
PRO-*who*-VP	*It wasn"t [me who pushed them over].*	VALIAN	2;7;21
N-*whose*-NP-VP	*And a brother called [Ross whose nickname was Spike]!*	MACWHINNEY	4;10;12

N-*which*-NP-VP	*He saw [some balloons which a man was selling].*	FLETCHER-SEVEN YEAR OLDS	6;11;9
PRO-*that*-NP-VP	*I mean [one that you had].*	LARA	3;3;22
PRO-*when*-NP-VP	*I hate [it when you say sister].*	MACWHINNEY	6;7;1
PRO-*that*-VP	*And [this that goes up there up there].*	WEISMER	3;8
N-*which*-VP	*(…) it annoys me actually xxxxx xxxx xxx xxx xxx and [a box which was so big] he did doesn''t know what (…)*	FORRESTER	3;0
N-*where*-NP-VP	*Theyre in [the bathtub where we played one night].*	INKELAS	2;3;5
N-*when*-NP-VP	*And [tomorrow when we do it again].*	MANCHESTER	2;9;6
N-*who*-VP	*Big bad wolf at [the big bad wolf who''s been in my bed]?*	BELFAST	2;7;27
N-*that*-NP-VP	*Not de cranberry juice [the cranberries that you bought].*	BROWN	3;8;26
N-*that*-VP	*Well that means that [the thing that''s coming out of your bottom].*	FLETCHER-FIVE YEAR OLDS	5;2;22

The Spearman's rank correlation shows very high values for inter-group production rates of the observed relative patterns (N=18). The degree of similarity in terms of how frequently each of them are used is both high and significant (see Table 30). The correlation value is the highest between the 4-6 and the adult age group, but the other two values are also extremely high, indicating that inter-group differences are barely existent as far as the production rates of the observed patterns are concerned.

Table 30. Spearman's rank correlation of the observed relative patterns across age groups

	0-3	**4-6**	**18+**
0-3	**1**	**0.919****	**0.907****
4-6		**1**	**0.969****
18+			**1**

**ρ=0,475 p=0.05 **ρ=0,625 p=0.01*

Altogether, five most frequent relative patterns take up 95-96% of all relative expressions in all age groups, which agrees with similar numbers surrounding most frequent items and expressions both in ditransitive and causative constructions. By far, the most frequent relative pattern in all three age groups is N-*that*-VP (e.g. *remember puzzles that are colored*; WEIST, 2; 10); when the frequencies are normalised with respect to the number of tokens, it

becomes evident that it is the most frequent pattern of relative clauses in all age groups. It is a construction where the object or the subject of the main clause becomes the subject of the subordinate clause (e.g. *The boy that plays football also loves basketball*). It is not surprising that this is the most common construction in all age groups, given that it represents a subject relative clause (which are processed more easily than object relative clauses) and makes use of the relativizer *that* which, unlike others, can relativize both animate and inanimate objects. The discrepancy between N-*that*-VP and other patterns is more pronounced in the two younger age groups than in adults, which to some degree resonates with 'item-based' assertions regarding the character of early language; i.e. this is confirmed only in part because the most salient frequency for the pattern is observed in the 4-6 age group (rather than the 0-3), where the pattern takes up approximately 44% of all relative constructions. The results may bear implications concerning the influence of input or the pattern-related processing difficulties (or both). It may be the case that children are additionally encouraged to (re)produce the pattern most frequently heard to an even greater degree, just as it may be the case that, at least in terms of ratios, they prefer it more because of the ease of production in comparison to others.

Table 31. Most frequent relative patterns (relative frequencies normalised per million tokens)

	0-3		**4-6**		**18+**	
N-*that*-VP	90	41%	484	44%	341	36%
N-*that*-NP-VP	49	23%	304	28%	280	30%
N-*who/whom*-VP	40	18%	147	13%	129	14%
N-*when*-NP-VP	14	6%	65	6%	76	8%
N-*where*-NP-VP	14	6%	45	4%	47	5%
N-*which*-VP	1	1%	7	1%	23	2%
	96%		96%		95%	

In order to put this into context, we may further observe the differences between the first and second most represented pattern in all three age groups, which is N-*that*-NP-VP (*I talked something that I knew one thing*; CLARK, 3; 1). While the frequency trends among the most common constructions obviously coincide between age groups with some minor differences in ratios, it should be noted that the drop in frequency between the first two categories (N-*that*-VP and N-*that*-NP-VP) is significantly lower in adults (36% →30%), i.e. the drop is a slightly more drastic in the 0-3 and the 4-6 age groups between the two most frequent patterns (41% →23% and 44% →28%, respectively), than it is in the adult group. Here, we can more easily observe the competition between the effects of frequency and the

difficulty of processing; despite the fact that this is still the second most frequent pattern in the younger age groups, the fact that the drop between the two is slightly more pronounced might indicate the role of processing difficulty in the production of relative construction NP-*that*-NP-VP. In other words, adults might be unburdened by this difficulty in use as opposed to children, who resort to alternatives more easily processed in terms of both cognitive and linguistic disentanglement.

Do these numbers suggest that the construction N-*that*-VP proves "easier" for children than the construction N-*that*-NP-VP? The difference between these two is only 6% in overall ratio of usage in the adult group, and in the younger groups the number amounts to 18% (0-3) and 16% (4-6). In other words, if we compared just the first two constructions, the difference between the first two is only 10% in the adult group, while in younger groups the difference is considerably greater: 30% in 0-3 age group and 22% in 4-6 age group. As far as ratios are concerned, it appears that for every N-*that*-NP-VP, there are 1.22 N-*that*-VP in the adult group, 1.59 in the 4-6 age group, and 1.84 of them in the adult group. Despite the fact that N-*that*-NP-VP patterns are almost as frequent in children's language environment as N-*that*-VP, children evidently still prefer the latter type. One major reason for this may be in the fact that the latter is also a construction in which the subject or object of the main clause also takes the role of the subject in the relative clause. When an additional argument is introduced into the given construction (like in N-*that*-NP-VP), the matter often gets complicated because the new nominal group tends to introduce a new (thus far unmentioned) subject of the relative clause. This could explain the greater difference in the frequencies of the aforementioned constructions among the younger groups (e.g. *the sauce that I dipped um that I dipped my egg rolls in*; HSLLD, 5;3)

The third most frequent pattern N-*who/whom*-VP[79] is slightly more represented on average in 0-3 age group (18%) than in 4-6 age group (13%) or in adults (14%). The manifestation of this pattern implies the relativized nominal element to represent an animate being, inviting the question of whether this fact plays a role in the production of relative

[79] Although the pattern has been dubbed N-*who/whom*-VP, aiming to initially include both relativizers, the overall number of instances with the use of the relativizer *whom* (one case observed in 4-6 and one in 18+ age group) was extremely low and thus basically immaterial to the research. Generally speaking, the relativizer *whom* is not commonly used outside of some formal or academic contexts, and the low frequency is even more accentuated in spoken language (as opposed to the written one). For example, from the fear of being prescriptively wrong or for the sake of simplicity, speakers tend to avoid using forms of words they may render as 'affected', such as the objective case in *whom*, and instead opt for *that*, no relativizer at all (zero) or even *who* (see Guy & Bayley, 1995).

construction in the youngest age groups. While it is ambitious to speculate about the direct role of animacy in the overall production of relatives, it is safe to ascertain that animate nouns are generally not abstract in nature, which surely plays a role in early language development. In other words, from early on children focus on items that are observable in their immediate surroundings such as concrete objects and people, rather than abstract entities. In this context, a slightly more frequent production of N-*who/whom*-VP is not unforeseen.

The difference observed between subject relative and object relative clauses across age groups is evident and statistically significant, χ^2 $(2, N = 2) = 8.6776, p < 0.05$, providing important insight in the degree to which the two alternatives occur (Table 32). The reason for this is in the fact that it shows the relevance of input but additionally also reflects the difficulty of processing that becomes pronounced in the younger age groups. In the 0-3 age groups, the portion of subject relative clauses takes up 62.5% of all relatives produced, whereas in the 4-6 age group it amounts to 59.8%, and in adults to 54.1%. In other words, the older the speakers become, the more frequently they produce object relative clauses. These results are in line with previous findings in the domain of language acquisition (Reali & Christiansen, 2007; Diessel, 2009a; O'Grady, 2011 etc.) and they directly relate to studies concerning the processing difficulty of the two relative variants which have consistently shown that object relative clauses require more 'processing power' than subject relative clauses (Holmes & O'Regan, 1981; Ford, 1983; King & Just, 1991; Gordon, Hendrick, & Johnson, 2001; Traxler, Morris, & Seely, 2002)

Table 32. Difference in the frequency between subject relative and object relative clauses

	0-3		**4-6**		**18+**	
subject relative clauses	135	62.5%	652	59.8%	509	54.1%
object relative clauses	81	37.5%	437	40.1%	429	45.6%

The extrapolated data from this comparative analysis of language used by the three studied age groups corroborates the findings of previous studies that explored the processing difficulties of relative sentences depending on subject/object positions in the subordinate clauses. The sentences where the object or the subject of the main clause becomes the patient/goal/instrument in the subordinate clause appear to require more effort in terms of processing and/or production. Syntactically speaking, the object relative clause construction entails an increase in complexity because it adds a nominal element which carries out the action expressed in the subordinate clause. The relative equation becomes more elaborate as it requires

of its speaker the cognitive effort to envision the event with more participants and at the same time presents a greater challenge in terms of the sole production of the utterance at hand, i.e. the introduction of an additional element in the construction to be uttered presents a more difficult operation from the standpoint of language performance.

Relating to the findings of this research in view of the interrelationship between processing difficulty and input is the study by Reali and Christiansen (2007); in their analysis of the American National Corpus, they found that pronominal object relative clauses (e.g. *the man that we assumed was lying*) are used significantly less than pronominal subject relative clauses (e.g. *the man that assumed we were lying*), with the ratio of nearly 2:1 in favour of subject relatives, but more importantly, following the corpus exploration they conducted self-paced moving-window reading experiments assessing the difference in processing difficulty between the two relative types with respect to exposure. Their conclusion was that pronominal object relative clauses were processed more easily when they were highly frequent, and that the speaker's exposure to a particular construction and the statistical information attained play a substantial role in the reduction of processing difficulty. In this context, it is interesting to interpret the minor but steady rise in the use of object relative clauses with the progression of age. If the results solely reflected the processing difficulty of the two constructions without other potentially mitigating factors such as the frequency of exposure, the data obtained would probably show greater discrepancy between the two constructions in the youngest age group in favour of subject relative clauses. Indeed, while it is true that the data does indicate a drop in object relative use with age declining (they are used 8% less in the 0-3 age group than in adults), the ratio between the two patterns in the 0-3 age group resembles the findings of Reali and Christiansen (2007), indicating that the overall production of object relative clauses is not that low (especially when the specificity of the corpus and the age group is considered). To sum up, the rise in the use of object relative clauses with the progression of age reflects the difficulty of processing between the two constructions confirmed by a number of studies, but the fact that they are still relatively frequent in the 0-3 age group might suggests that the role of language input is sufficient to overcome these difficulties very early on.

While many studies have delved into developmental and processing-related aspects of the acquisition of relatives with regards to various types of relative clauses and relativizers, few have looked into the distribution of the relativized elements themselves. This partly relates to the fact that the discussion on the types of noun phrases that are being relativised fall beyond

the scope of relative clause-descriptions and belong to the realm of relative constructions. Beside the animacy aspect, little can be said on the type of the relativised elements. In terms of parts of speech, the expectation is almost always that the relativised element will be an isolated or complex noun, much more rarely replaced by a pronoun. But exactly how dominant are the nominals as opposed to pronouns in child and adult speech? In this thesis, the analysis of the relativised elements in relative constructions looked at the difference between nominal and pronominal relativisation (Table 33). The results show a huge pronoun-noun discrepancy in the position of the relativised element, with more than 96% of relativised noun phrases being occupied by an isolated or complex nominal across all three age groups.

Table 33. Difference in the frequency between relativized nouns and pronouns

	0-3		**4-6**		**18+**	
PRO-REL-(NP)-VP	8	3.9%	27	2.5%	36	3.8%
N/MOD-N-REL-(NP)-VP	208	96.3%	1063	97.5%	906	96.3%

What is the implication of this result and why do the pronouns tend to be so underrepresented as relativized elements? More precisely, why are they so underrepresented in the younger age groups, especially considering that children do not shy away from the use of pronouns when it comes to other constructions we observed? In other words, there is no reason for such misbalance considering that there is nothing ungrammatical about the use of pronouns in the place of a relativized element (sentences such as *it was him that had done it* or *it was there where it happened* are completely acceptable, if not highly regular).[80] Nevertheless, the case seems to be that adults truly produce these constructions rarely in comparison to constructions with nouns as relativised elements. There is no point of reference with which to compare the data on relativised elements in adults, but there is also no obvious reason why adults would not use nouns more frequently than children, particularly because they have more lexical items at their disposal to refer to target elements on the account of their vocabulary being more developed. And yet, the data clearly shows an overlap in terms of frequency when it comes to the use of pronouns in the relativised element positions. One of the explanations for this may be in the fact that children indeed rely on the input to such extent that the production frequency

[80] Nonetheless, it is predominantly non-restrictive relative clauses that tend to be headed by the pronoun because the pronoun in itself is essentially specific and definite, whereas the head of the restrictive relative construction is definite and specific only because of the relative clause.

of various syntactic patterns greatly resembles the input received from adults. In this context, the statement ‘if children rarely hear something, they rarely use it’ aligns well with the data obtained and represents the usage-based hypothesis in a nutshell. Such results and inferences consistently persist across each of the analyses presented thus far.

5.2.3. Lexical arrangement and overlap

The subsequent aspect of similarity between age groups in relation to the formation of relative constructions involves examining the particular vocabulary employed in various construction slots. Particularly intriguing in the context of inter-age group correspondence is the inquiry into whether there exist discernible patterns based on specific items regarding the selection of relativizers and the nominal constituents that head the relative construction.

As far as the relativizer slot is concerned, the results (Table 34) clearly demonstrate that the relativizer *that* is the most frequent choice in all age groups, taking up at least 2/3 of the overall production of relatives. It means that the speakers, whether child or adult, chose to use the relativizer *that* in 6 to 7 instances out of 10. The second most frequent choice (again for all three age groups) is the relativizer *who*. When it comes to the choice of the relativizer, the frequency trends are very similar in all three age groups, i.e. it is almost identical, but it should still be noted that the share of certain relativizers differs considerably. For example, in younger groups, there is a slightly higher concentration of the first two types of relativizers than in the group of adults. In the youngest group, the first two relativizers make up almost 86% of the total production of relative sentences, in the 4-6 age group around 88%, while in the adult group this figure is approximately 82-83%. On the other hand, compared to other age groups, the results demonstrate the greatest dispersion in the adult one. Albeit only slightly, the types of relativizers employed by adults are less concentrated among the first few, and there seems to be some observable increase of use of those otherwise less frequently represented. Although slight differences in dispersion should be taken ‘with a grain of salt’, it is probably not a coincidence that the dispersal is more balanced precisely in the adult age group regardless of the fact that it is characterized by child-directed speech. The finding is even more interesting in view of the fact that the observed relative constructions do not prove to be the most frequent in the adult group (the adults might perceive relative constructions too complex to convey a

particular idea, thus resorting to alternative expressions in their child-directed speech). In this context, regardless of the fact that the adults might be more restrained in their use of relative constructions (hence the lower overall frequency of use than in the age group 4-6), they still demonstrate a greater diversity in their production. Thus, a comparative analysis of the frequency of use of certain relativizers in relative constructions reveals the following: the age groups correspond in the frequency order in the use of particular relativizers, but the variety of their use is somewhat more balanced only at a later stage of language development.

Table 34. Relativizers across age groups (ratios)

	0-3	**4-6**	**18+**
that	66.5%	74.0%	68.2%
who	19.1%	14.1%	14.3%
when	6.4%	6.4%	9.1%
where	6.9%	4.2%	5.0%
which	1.0%	1.2%	3.0%
whose	0.1%	0.2%	0.4%

The data obtained on relative constructions and the most frequent relativizers is in line with the usage-based interpretations of language acquisition process. As predicted by usage-based claims, the linguistic input that children receive is crucial for the development of spoken language, but more importantly, during the process of language acquisition, children seem to be relying partly on imitation (Tomasello, 2000a).[81] In other words, if in some cases the relativizer *which* was the most common relativizer in the speech of adults, we would expect the same in the speech of children. If the language environment plays a key role in acquisition, the distribution of relativizers in the youngest groups should mirror the distribution of relativizers in adult groups.[82]

In fact, the only difference frequency-wise can be observed in the use of the relativizer *where*, i.e. although a fairly small, the variation in the use of *where* and *when* reveals a

[81] Here, it needs to be noted that Tomasello does not use the term 'imitation' literally, but refers to imitation in the context of learning (the so-called 'imitative learning'), which implies that the child can understand the intention of the speaker ('intention-reading') in addition to their ability to reproduce the linguistic input.

[82] On the other hand, nativist arguments should find support in any deviations from the input; the lexicon is seen as something naturally attained in the context of their immediate linguistic environment, but the children's selection of relativizers going beyond the role of input frequency (especially because children are not as burdened by the pragmatic aspects of conversation) would be an indication of less conservative language use by children. Thus, a somewhat increased diversification of the use of certain relativizers in adult speakers is expected according to usage-based approaches, but not necessarily according to nativist approaches.

preference towards *where* in the youngest group, whereas the situation is reversed in other age groups. Quite possibly, this information might suggest something about the cognitive development of children – one may speculate that the concept of time for younger children (although attainable) is more difficult to master than the concept of place. One evidence to support this is also the fact that time tends to be expressed as a spatial metaphor, meaning that much of the child's learning of language conforms to their knowledge of the space surrounding them (see Clark, 1973).[83] It thus follows that children develop the conceptualization of time at the later point than that of space, but what this data undoubtedly confirms is that – at one point, when children begin to grasp the concept of time more readily – the share of relative clauses that supplement sentences with temporal content increases compared to those that relativize locational referents.

What is more, the raw frequency of the relativizer *which* amounted to only 8 times in a relative construction (where the referent was an explicit nominal group) in the 0-3 age group (out of total 673 times that the lemma *which* appears in the 0-3 CHILDES corpora). The question is why the relativizer *which* is so demanding for children to learn, especially if we take into account the fact that it semantically, contextually, and to a large extent syntactically corresponds to the most common relativizer *that*.[84] It is true that the frequency of the relativizer *which* is relatively low in the adult group, but it still occupies a significantly larger share in the spoken language of adults (even when it comes to speech aimed at children) when compared to younger age groups. Such results favor the item-based claims, propagated by advocates of the usage-based approach, which assume that the process of acquiring certain constructions rests on the initial production that is concentrated around a few of the most accessible and simplest lexemes that later lead to further development.

After looking at the relativized nouns in the pattern N-*that*-NP-V (see the end of discussion in Section 5.2.2), it again becomes clear that there is an astounding amount of

[83] Moreover, Haspelmath (1997) shows that most languages issue the lexical and grammatical material for expressing time from the already available expressions for defining space relations. In fact, according to him, this is a language universal as there do not seem to be languages departing from this trend.

[84] Naturally, there are significant differences in the use of *which* and *that* in the sense that the former is more constrained than the latter as it is restricted to inanimates, whereas *who(m)* are used its counterpart for animates. The point, however, is that it can stand in the place of *that* whenever it relativizes an inanimate object. Moreover, *which* tends to be more syntactically flexible as it can be used in non-restrictive relative clauses, and it can relativize entire clauses instead of solely relativizing NPs (e.g. *He only loves to play chess, which is a sign of a madman*), whereas this is not viable for *that*, all of which should speak against such low production rates for the relativizer *which*.

similarity with regards to what is being relativized (see Table 35). There is no obvious reason for great similarities in the relativized slot given the range of nouns that can occupy it and a relatively restricted number of constraints on the slot (if any) that would reduce the said range, and yet, several nouns stand out as the most common ones in all three age groups. First, the nouns *thing* and *something* take the first two spots in all three age groups. Both terms can stand to represent a number of things, so the reason for their overproduction in the younger groups is twofold: the breadth of items they refer to depending on the context, and the frequency of their relativized slot representation in the input. Similar case is with the nouns *one*[85] and *stuff*, which can again stand to replace a number of referents in the surroundings. What is more, the nouns such as *toy(s)* and *book(s)* appear across all three age groups regardless of being more specific than the aforementioned terms, and there are nouns such as *car(s)*, *ball(s)* or *time(s)* which rank highly in the input and are either found in 0-3 or 4-6 age group among the most frequent items.

Table 35. Relativized nouns in the pattern N-that-NP-V

	0-3	**4-6**	**18+**
1.	*thing(s)*	*thing(s)*	*thing(s)*
2.	*something*	*something*	*something*
3.	*toy(s)*	*stuff*	*book(s)*
4.	*ones*	*everything*	*toy(s)*
5.	*book*	*ones*	*car(s)*
6.	*beans*	*time*	*box(es)*
7.	*movie*	*book(s)*	*one(s)*
8.	*car*	*ball(s)*	*ball(s)*
9.	*food*	*song(s)*	*time(s)*
10.	*table*	*guy(s)*	*picture(s)*
11.	*part*	*way(s)*	*house(es)*
12.	*stuff*	*game(s)*	*song(s)*
13.	*guy*	*toy(s)*	*stuff*

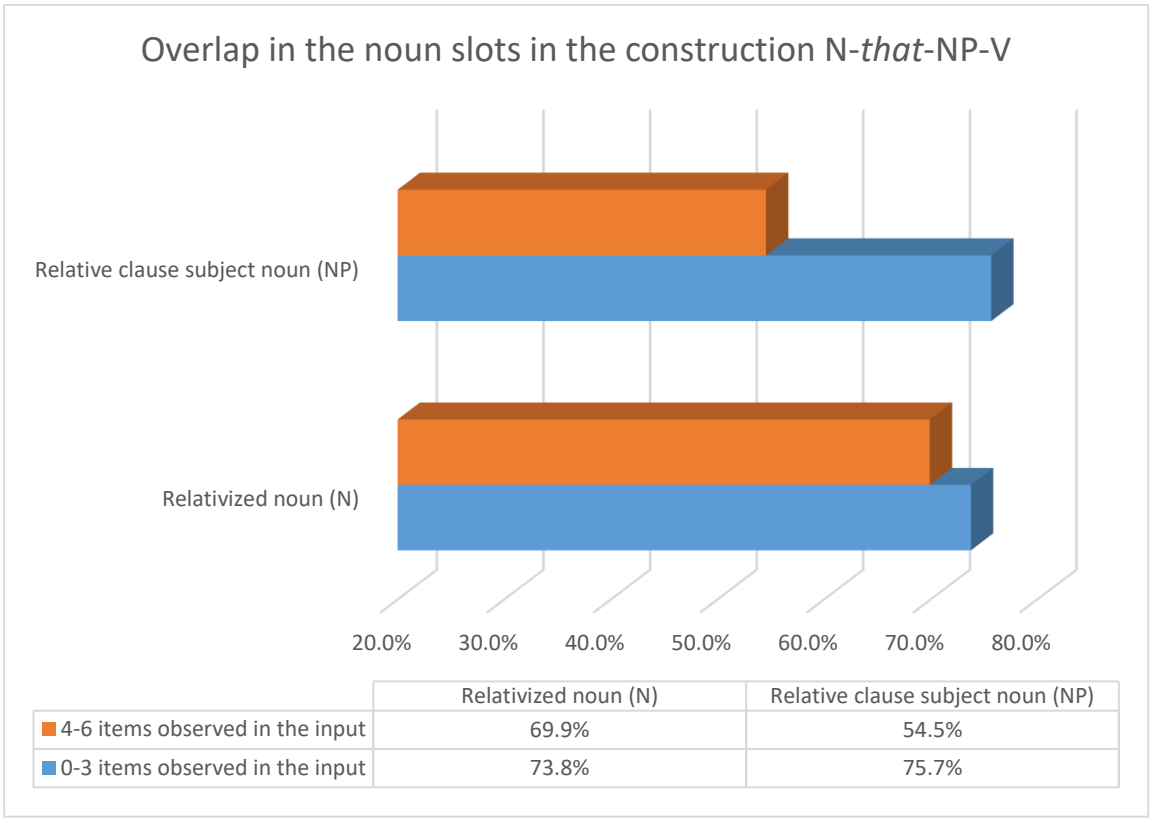
It also seems that the youngest age group is more restricted to concrete objects as opposed to abstract ones; for example, the noun *time* ranks highly in the 4-6 age group, and it does not appear in the youngest age group. Even the noun *song(s)*, which does not physically manifest itself in the environment appears highly ranked in the input and the 4-6 age group, but not in the 0-3 age group. This limitation is one that goes beyond the input frequency in the earliest speech production and relates to the abstractness of concepts referred to by the speakers.

[85] Similarly to *something*, *one* was recognized as a noun by the Sketch Engine tool, and not a pronoun.

As such, the abstractness of concepts plays a role in the earliest language development and in part confirms the role of general cognitive mechanisms accounting for lexical acquisition (e.g. children's tracking of adults' gaze in order to determine the referents of language produced by said adults).

In order to better determine the extent to which the young age groups correspond to adult language use when it comes to items used in specific construction slots, it is best to look for the percentage of items that have been found to match between the observed age groups (see Figure 12). Again, the proportion of items that has been observed in the adult speech is higher in the 0-3 age group than in the 4-6 age group. The discrepancy is more evident when it comes to subject nouns of the relative clause than when it comes to relativized nouns, but the results in both categories agree with the hypothesis that the youngest age groups would be more anchored to lexical items used in the adult speech when compared to their slightly older counterparts.

Figure 12. The proportion of lexical items observed in the input



	Relativized noun (N)	Relative clause subject noun (NP)
4-6 items observed in the input	69.9%	54.5%
0-3 items observed in the input	73.8%	75.7%

Moreover, the degree to which this happens is relatively high, given that for each of the two categories more than 70% of the items used in the specific construction slot have also been

observed in the adult group in the same slot. If we assume that the discrepancy is a reflection of the children's reliance on input, we must question why there is a greater 'imitative tendency' in the relative clause subject noun position than in relativized nouns; that is, why is there a greater diversity on the part of older children when it comes to the subject slot. First, the relative construction is a complex one which requires a great deal of processing power to comprehend and produce, and object relative constructions have proven to be more challenging than subject relative ones in that regard (see Section 5.1.3). It is only logical to assume that children will then initially stick to the input as much as they can, as going beyond it might impede their capacity to process and produce the said constructions. As they grow older, they start to be particularly flexible with the relative clause subject noun, which needs not be particularly surprising given that the relative clause represents a sentence in itself, i.e. the position of the subject is consequently one of the least constrained syntactic slots in the language – hence the rise in the independence from the input. Another aspect to consider is the difference in the number of constructions observed in each of the age groups (in the 4-6 age group the final raw frequency was 509, whereas in the 0-3 age group it was 214), which affects the number of different types in each of the construction slots (in the 0-3 age group there were 124 types of relativized nouns and 37 types of relative clause subject NPs; in the 4-6 age group there were 226 types of relativized nouns and 55 types of relative clause subject NPs).[86] Although this is certainly a contributing factor to the observed discrepancy between the 0-3 and the 4-6 age group in terms of their similarity to the input, it would be a mistake to treat it as an only one; i.e. consider the fact that there is a similar degree of overlap between the 0-3 age group and the adult group in both slots (73.8% and 75.7%), regardless of the significant difference in the number of types observed in each of them (124 and 37), meaning that an analogous portion of them have been observed in the input despite lower probability of that being the case.

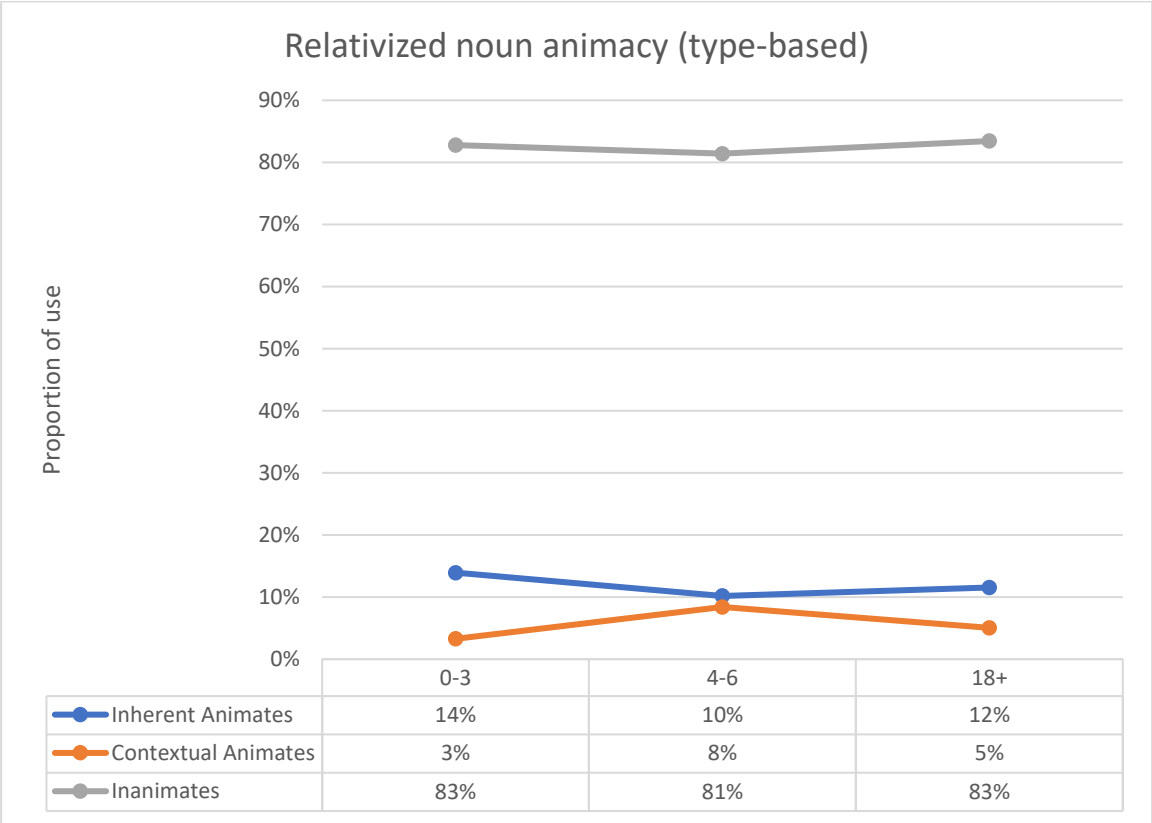
[86] As far as the adult group is concerned, the number of different types in the relativized slot position was 1370, and the number of different types in the relative clause subject NP slot was 408.

5.2.4. Animacy

In relative constructions, the relativized slot may be worthwhile to observe in view of language development, as it tells us what exactly the speaker wants to elaborate upon. The relative clause typically serves to provide an additional information on the relativized item in the form of a subordinate clause, and requires of a child to understand and apply a number of linguistic concepts when producing it. For example, the relativized noun calls for the use of different relative pronouns, which children must learn to use appropriately. In turn, this requires an understanding of the grammatical role of the noun being modified and its relationship to the rest of the sentence. In the context of comparison between age groups, the animacy of the referent noun phrase in the relativized construction may reveal differences with respect to potential preferences concerning the character of the noun that the speaker is referring to. In other words, it is possible that children may differ from adults in the degree to which they prefer to relativize animate and inanimate nouns, which could then potentially inform us about the developmental character of their language use.

For the same reason, it is best to provide an analysis of the relative construction N-*that*-NP-V, where the relativizer takes both inanimate and animate nouns. When it comes to the animacy of the relativized items in the aforementioned relative construction, the data indicates a significant prevalence of inanimate nouns being relativized across all three age groups. On average, it turned out that inanimate nouns were typically the choice of all ages in at least 4 out of 5 sentences produced. The second most frequent choice were the inherent animates, which cover nouns generally considered as animate (living beings) and proper nouns (names of people). The third most frequent choice were the contextual animates in all age groups, although the difference was less obvious in the 4-6 age group, as opposed to the 0-3 age group and adults.

Figure 13. Relativized noun animacy (type-based)



One of the reasons for the possible prevalence of inanimate nouns in the position of the relativized slot may be attributed to the fact that there are generally less animate nouns than there are inanimate ones in the English language, and given that the relativized slot in the aforementioned construction is open and flexible to different types of nouns meaning-wise – the frequency argument appears legitimate. However, this is not the case for the data in question. In fact, the numbers show an even greater discrepancy between the category of inanimates and the rest when the calculation takes into consideration frequency of production (and not just diversity), which would be unexpected if the previous hypothesis were true (in other words, regardless of a more restricted number of animate nouns in general, the repeated use of the same in the position of the relativized item would tilt the scales in favour of the category of animates (whether contextual or inherent).

A more fine-grained look at the animacy of the relativized noun in the construction N-*that*-NP-V shows that the second position in all age groups is occupied by general animate nouns (as opposed to names for instance, which generally group together with typical animates to form the 'inherent animates' category). Especially interesting is the fact that children much

less prefer to use names as relativized nouns and instead resort to general animates (0-3 age group) or animate toys (4-6 age group). Part of the reason may be in the relativizer *that*, which competes with *who* when relativizing general animates, and especially names of the people. Since children encounter *who* in the input more frequently than *that* with names of people in the position of the relativized slot, it is not surprising that they opt for other types of animates/inanimates in the said construction.

Table 36. Relativized noun animacy (type-based)

	0-3	**4-6**	**18+**
INAN	83%	81%	83%
ANc	13%	9%	7%
ANp	1%	1%	4%
ANt	2%	7%	4%
ANb	1%	1%	1%

The cross-age group resemblance in the animacy of the relativized slot is particularly interesting considering the great degree of freedom in terms of choosing which noun to relativize (as far as nouns are concerned, the potential choices are always greater than with verbs).[87] But even in terms of potential noun diversity, we would still expect the number of semantic constraints on the choice of the relativized noun to be smaller than the number of constraints on, for example, indirect objects in the ditransitive construction (or even direct ones for that matter). There is no reason to predict such great disparity between inanimates and animates in the position of the relativized slot.

The results are in line with usage-based predictions of early language character, i.e. the saturation of the relativized noun's animate character around the first two types (primarily inanimates, and then general animates) is highest in the youngest age group (96% in total), while the adult age group shows a more even distribution between the types when compared to the younger ones. The data is consistent with the predictions that early language would resemble the adult one (especially in relation to most frequent items), while also leaving room for the greater flexibility of adults evident in the range of items used in a particular slot. The fact that the data on animacy shows similar tendencies as data on particular items and patterns (in view

[87] For instance, the selection of items that can be relativized is considerably greater than the number of causative verbs that can be employed in the periphrastic realization.

of cross-age correspondence) tells us that the item-based character of the early speech goes beyond the reproduction of the items used in the input, and reflects on the character of these items as well. In other words, although children learn language 'imitatively' to a certain degree, they also pick up on the semantic and syntactic constraints of these items in the process, and tend to employ items which share the features of those encountered in the input in the target construction slots.

5.2.5. Lexical richness

A more detailed review of the construction N-*that*-NP-V reveals a greater jump in the utterance length between the 0-3 and the 4-6 age group, while the disparity between the latter and the adult group remains much smaller. On average, the 4-6 age group tends to use as much as three morphemes more per utterance in comparison to the 0-3 age group. This tells us that, despite mastering the fundamentals of the construction, it is only after the children turn four years of age that they begin to use the construction in more elaborate ways. Given that the MLU stated in Table 37 is a reflection of the utterance, rather than the construction in isolation, it is safe to assume that the children's learning trajectory is initially characterized by the production of integral elements of the relative construction (in object relatives represented by: the relativized item, the relativizer, the subject of the relative clause and the verb of the relative clause), and only later by the elaboration of the linguistic context in which it is produced, such as learning to embed the construction into the relative sentence and beyond.

Table 37. General overview of the sentences containing N-that-NP-V across age groups

	0-3	**4-6**	**18+**
Token count	2132	6513	71902
Word type count	431	855	3626
Character count (without spaces)	8552	24952	295103
Utterance count	206	497	5251
Open class words	935	2816	31498
Closed class words	1167	3571	39437
MLU (morpheme-based)	11,29	14,09	14,78

The indices of LD between the three age groups in utterances with the relative N-*that*-NP-VP constructions remain relatively stable. The same case has been observed in both

ditransitives and causatives; the numbers differ between the three constructions, but remain relatively stable across age groups when it comes to each particular one. When it comes to utterances containing the aforementioned relative pattern, the index does not go beyond 0.42 (as in the 0-3 age group). This means that all age groups, when producing utterances revolving around the relative N-*that*-NP-VP, approximately produce 4 content words per 10 tokens on average, which does not suggest particularly 'dense' utterances on the part of any age group. The relative construction in itself always implies some function words such as relativizers, but the overall density of the pattern is subject to changes depending on the number of articles and the like. The relative constructions are derived from key elements of relative sentences, and being embedded in such larger structures allows for the introduction of various function words such as conjunctions, prepositions etc. Nonetheless, the LD values do not seem to reflect any prowess on behalf of older children or adults in this regard, nor do they reveal any differences across age groups beyond values that could arguably be attributed to chance. In order to obtain a more detailed introspect, it is necessary to see the behaviour of other indices of sophistication and diversity.

Table 38. Lexical richness for the relative N-that-NP-VP

	Lexical density and sophistication			**Lexical diversity: range and variation**				
	LD	LS1	CVS1	NDW	NDW-ER50	MSTTR-50	RTTR	SVV1
0-3	0.42	0.24	0.69	412	35.4	0.64	8.85	26.79
4-6	0.4	0.2	1	470	36	0.66	9.14	28.74
18+	0.41	0.21	0.77	604	36.1	0.7	10.99	30.94

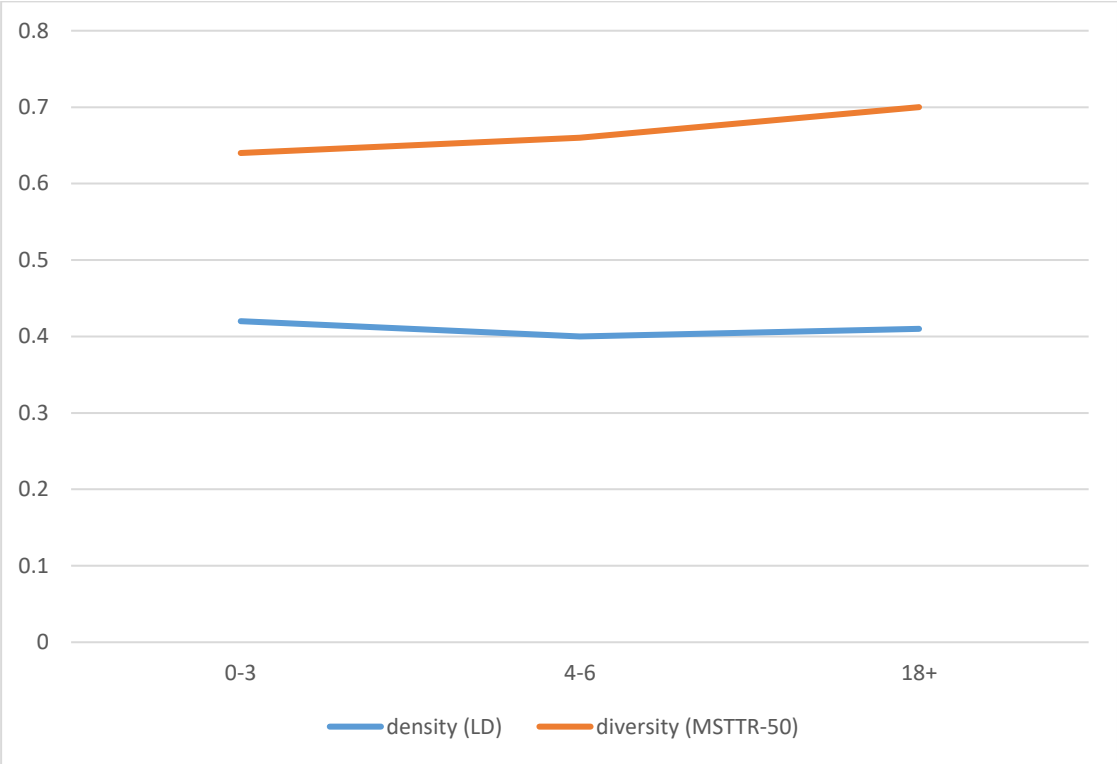
As far as lexical sophistication is concerned, the data on LS1 oddly indicate a drop in the values between 0-3 and 4-6 age groups, which then share similar values to adults. This is not expected and requires further examination into the forms of lexical items marked as less frequent; it may be the case that the lexemes truly reflect the drop in the use of less frequent items in general, but it is also possible that some of the items were wrongly designated as 'sophisticated' due to their infrequency[88] when in fact they represented nonce words. In terms of verb sophistication, the CVS1 index shows a separate trend from LS1. Here, the lowest score

[88] Due to the way in which sophisticated words tend to be calculated, some very frequent words in child language could potentially belong to the group of "sophisticated" items, as they can be very rare in corpora but frequent in everyday usage (e.g. *fork*, *spoon*, etc.). The measure used in this research counted lexical items as sophisticated if they were not on the list of the 2,000 most frequent words in the British National Corpus (see Lu, 2012, p. 192).

is found in the 0-3 age group, followed by adults, and finally by 4-6 age group. In other words, in the utterances containing the relative N-*that*-NP-VP construction, the least frequent verbs are found in 4-6 age group. Although this is not necessarily to be expected, there are reasons why such results bear certain legitimacy. Relative constructions are the most complex constructions out of those observed in this corpus-based study, and they may appear as such to the adult speaker who then instead takes particular note when producing them to simplify the construction and the surrounding elements (see discussion in Chapter 6, Section 6.1 as well). In this context, the child-directed character of child-directed speech may be more pronounced in the constructions that require more processing on behalf of speakers (and listeners in particular); if such reasoning is adopted, the lexical simplification implying the use of more frequent lexical and verbal items is not necessarily unanticipated. However, it also needs to be added that verb-variation values do not follow the same trends as those of sophistication; i.e. SVV1 shows a small but relatively stable increase across age groups, suggesting that the variety in the use of verbs in utterances containing the relative constructions indeed increases with age, but the choice of rarer verbs does not necessarily follow this trend.

The lexical diversity indices also show an increase across age groups, with MSTTR values ranging from 0.64 in the 0-3 age group, 0.66 in the 4-6 age group and 0.7 in adults (Figure 14). The increase is similarly proportional for RTTR values, with both indicating higher degrees of lexical diversity with the progression of age. Given that the lexical density index is relatively stable, we can again assume that the progression in diversity does not interact with the ratios between lexical and grammatical words. Naturally, an increase in diversity is to be expected with the progression of age, but the question remains to which degree are we to expect it, especially in relation to child-directed speech. As already stated, relatives are presented as more intricate language patterns in comparison to other constructions observed by this research, and the question is whether the adult speakers take this fact into consideration when talking to their children and purposefully simplify the utterances containing these constructions. It is difficult to evaluate whether the measures of diversity speak against such suppositions, and whether this increase in diversity (regardless of the fact that it exists) is actually lower than expected. This is addressed later in the thesis in relation to other constructions, and at this point it suffices to notice that both indices of lexical diversity taken into account demonstrate an increase across age groups.

Figure 14. Density and diversity in the relative N-that-NP-VP



When it comes to the measures of variation, the differences across age groups are slightly more obvious with NDW than with NDW-ER50. In the 0-3 age group, children produce slightly over 400 different words per 200 utterances containing the relative construction, while in the 4-6 age group, this number increases to 470 words. The spike is even greater when we just look at the adult group, where the same 200 utterances contain approximately 600 different words. However, the same numbers are considerably reduced when we look at NDW-ER50, which leads us to believe that the greater number of different words observed in older age groups is a relative side-effect of the average utterance length (which is corroborated by the MLU values for the same samples). In other words, the data indicates that as the children grow older they tend to produce lengthier utterances (that contain the relative N-*that*-NP-VP construction) with greater variety of lexical items employed. However, when smaller subsamples are taken into consideration, this variety tends to reduce considerably, which leads us to believe that the variety of lexical items is primarily a side-effect of the increase in utterance-size. Although variety is a separate construct from construction and utterance length, the values of variety converge around the same numbers when the subsamples are balanced in token-length, meaning that we can assume either one of the two following claims: (1) older children and adults in child-directed context are not necessarily diverse in their use of utterances

containing relatives or (2) the youngest category of children exhibits surprising amounts of variety in such utterances. In conclusion, the spikes in some of the categories concerning utterance length, verb sophistication and even diversity to some degree can be observed, but if we take away the relatively rare production of relatives on the side of the youngest category of children – when they do produce such utterances, some of their density and variation indices compare relatively well with those of older children and adults.

5.3. In sum

- The initial examination of the data concerning relative constructions indicates that the youngest age group (0-3 years) exhibits the lowest overall production of relative constructions by far. The production in this age group is approximately five times lower compared to the 4-6 age group. The 4-6 age group demonstrates the highest frequency of relative constructions being produced, followed by slightly lower overall numbers in the adult group. This is primarily attributed to the fact that the adult subcorpora mostly consists of child-directed speech, which is why adults may resort to simplification of their language, especially when it comes to constructions which they themselves would perceive as complex (such as relatives). The substantial difference in the usage of relative clauses between the age group 4-6 and the age group 0-3 is indicative of the children's language developmental stages.
- Regarding the particular syntactic sub-patterns of relative constructions, there is once again a noteworthy similarity between age groups in terms of their production frequency. The constructions that are least frequently produced relativize pronouns, and they also tend to be object relative clauses, which are generally more challenging to process compared to subject relative clauses. Apart from this, there does not appear to be an apparent reason for their underrepresentation in the younger age groups, except for the fact that they are rarely encountered in the input.
- The findings clearly indicate that the most commonly chosen relativizer across all age groups is *that*, accounting for at least two-thirds of the overall production of relative constructions. The second most frequent choice among all three age groups is *who*. When it comes to the selection of relativizers, the frequency trends are similar across all age groups, though it should be noted that the distribution of certain relativizers

differs significantly. The analysis reveals that the age groups correspond in terms of the frequency order of relativizers, but the diversity of their usage becomes more balanced at a later stage of language development.

- As far as the overlap in nouns is concerned, the results clearly indicate that the greater proportion of items produced in specific slots by the 0-3 age group is found in the input in comparison to 4-6 age groups. The results go in line with the hypothesis that the earliest age group will be more restricted to (re)producing patterns heard in their language environment with respect to their lexical specificity. The first hypothesis is again confirmed on several levels. There is a striking similarity across age groups in terms of lexical items employed in specific construction slots, even when it comes to less constrained positions such as relativized nouns. While there is no clear indication of greater concentration of the most frequent lexical items in the youngest age groups, the results agree with the supposition that the 0-3 age groups would lean more heavily on the adult input – hence the greater degree of overlap.
- On average, it was discovered that inanimate nouns were consistently chosen in at least 4 out of 5 sentences produced by individuals of all age groups, which is surprising considering that there is no apparent reason to expect such a significant difference between inanimate and animate nouns in the position of the relativized slot. The second most frequent choice was inherent animates, followed by contextual animates in all age groups, although the difference was less pronounced in the 4-6 age group compared to the 0-3 age group and adults. The fact that the data on animacy exhibit similar patterns as the data on specific items and patterns (in terms of cross-age correspondence) might indicate that the item-based tendencies of early speech go beyond the mere replication of the items present in the input, reflecting the nature of said items as well. In other words, the data on animacy is not as revealing in relation to the hypothesis predicting more constrained lexical choices in terms of animacy in the youngest age groups because all age groups seem to be relatively restricted in the observed construction slot. This does not go against the hypothesis that children's distributional learning of lexical items eventuates into correspondence in production rates with respect to said items' features in the observed construction slots, but it also does not reveal more about the impact of most frequent items in comparison to others, with all being reproduced in similar degree.

- The indices of lexical diversity and complexity suggest that, if we take away the relatively rare production of relatives on the side of the youngest category of children (0-3), the lexical variation is surprisingly high; that is, when they do produce such utterances, some of their density and variation indices compare relatively well with those of older children and adults. There is some increase in the range of different words with regard to the production of relatives, but other indices remain relatively stable across age groups which goes against the hypothesis predicting progression in terms of lexical richness and complexity despite item-based similarity. This either suggests the moderated use of relatives on the side of adults (being aware of the construction's complexity they may consciously mitigate its elaborateness) or the fact that children—once reaching the ability to truly produce the relative construction are already showing significant advances with respect to said lexical intricacy.

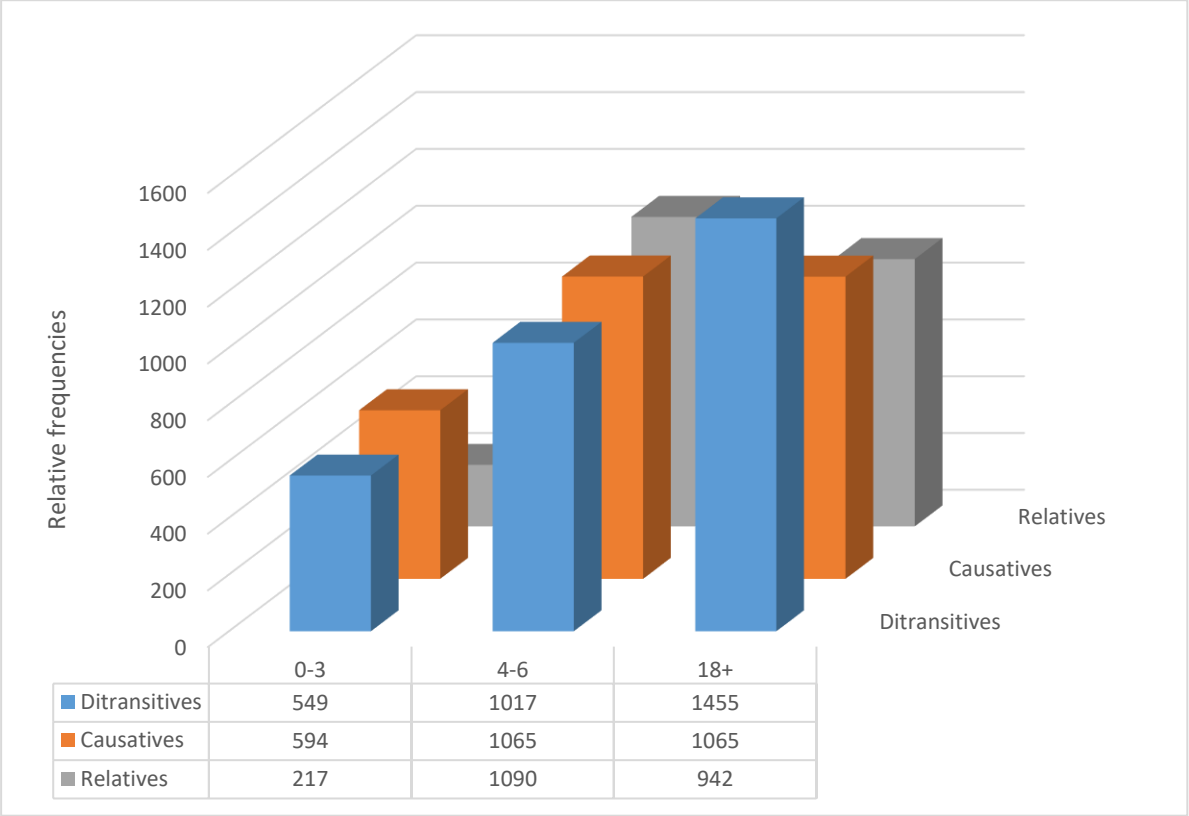
6. CONVERGENCE OF DATA AND FINAL DISCUSSION

6.1. Inter-constructional data

The initial look at the distribution of frequencies across age groups within certain constructions seemed to have shown shared tendencies in terms of direction; the age group with the lowest number of produced constructions was always the 0-3 age group in terms of relative frequencies, and when it comes to causatives and relatives, the 4-6 age group came pretty close to the adult group as far as the total frequency is concerned. This confirms the idea that, after the age of three, children have reached an almost mature degree of competence in their first language. However, some important differences become obvious when different constructions are observed simultaneously. In a sense, the cross-analysis of frequencies regarding the three constructions studied in the research shows unexpected results. It was initially hypothesized that the frequency of these syntactically different constructions in the younger age groups would reflect the overall complexity of the observed constructions (see Section 1.4 for the operationalization of the term). As it turns out, these assumptions turned out to be true only for the adult group, which is where the processing difficulty should affect the results the least. In other words, if the frequency reflected the processing difficulty of a particular construction, we might expect a steady decrease starting from ditransitives, followed by periphrastic causatives and finally by relatives (that is, if we consider relative constructions to be more difficult processing-wise than periphrastic causatives and ultimately ditransitives). Of course, there is a good explanation for this given that we are primarily talking about child-directed speech, and adults tend to simplify their language in such setting; it is quite probable that they adjust their language according to what they think would be easier to process for children.[89]

[89] While it may seem that this hypothesis has been somewhat taken for granted throughout the thesis, a number of different studies have already confirmed these suppositions about child-directed speech (see Pine, 1994; Cameron-Faulkner et al., 2003; Saxton, 2009 and others), including more recent ones done on some of the CHILDES subcorpora (see Jones et al., 2023). It would certainly be useful to compare the results with a corpus of adult (spoken) language, but this remains a task for further research.

Figure 15. Frequency of constructions across age groups (relative frequencies normalised per million tokens)



	0-3	4-6	18+
Ditransitives	549	1017	1455
Causatives	594	1065	1065
Relatives	217	1090	942

From a methodological standpoint, one could justifiably ask the question as to why the frequencies of these constructions can even be comparatively observed. Clearly, one construction can incorporate a wider range of phenomena and functions, i.e. when it comes to ditransitives, only the double-object dative was observed and had the analysis included it's prepositional alternation, chances are that the overall frequency would be considerably higher when compared to the other two construction within each age group. However, because there is data for each group separately and because the same extraction method was applied across age groups with individual constructions, it is possible to interpret the in-group differences by taking one group as an example of the expected distribution (e.g. the group of adults) and interpret the rest within the same context. For instance, if ditransitives are the most frequent construction in the adult group when compared to the rest, the same tendency should be expected in 0-3 age group. In some sense, we can use the adult group as a form of control group

to interpret the deviations from the norm in the inter-constructional patterns within 0-3 and 4-6 age groups.[90]

Perhaps, the most interesting results are found in the age group 4-6, where the production frequencies of all three constructions are relatively similar. What does this tell us about the interrelationship between productivity and complexity? First, we need to note the undeniable jump between the 0-3 age group and the 4-6 age group when it comes to the use of all three different constructions.[91] Clearly, the surge in use of the three constructions must be the result of the increase in the overall mastery in view of producing more complex utterances; i.e. we can safely assume that the ease with which these constructions are produced is significantly increased in the age group 4-6. The frequency of occurrence in and of itself is not an indication of language proficiency, but it directly reveals the rest of the language produced by each age group (given that we are speaking of relative frequencies normalised in relation to the size of the age specific subcorpora), in that sense indicating the character of each group's overall language variety, especially when corroborated with the analysis of three different constructions. In the context of relative frequencies, the fact that something has been produced infrequently means that something else was produced more frequently; but what does it mean when the same dynamic has been observed for three different constructions? While there is nothing groundbreaking in observing that the more simple utterances, as well as those incomplete or 'gibberish' in general, take up a larger part in the earliest language than any other, the data calculated for the three constructions gives us an idea of how big these ratios are (and what they might depend on).

Secondly, it appears that as the syntactic/grammatical mastery slowly moves towards its peak in the 4-6 age group, the production frequency of these constructions no longer depends on their levels of syntactic complexity. In fact, we can assume that semantic and pragmatic issues become more relevant in this group, as opposed to the syntactic ones with which they struggle in the beginning. In the 0-3 age group, there is clearly a 'competition' between

[90] This is by no means an ideal approach to data interpretation, especially if we consider the fact that child-directed speech tends to be very different from other spontaneous adult speech (as noted several times in the thesis). However, the analysed speech still constitutes an input that most probably presents itself as the first language (re)production objective to children in the study, and is perhaps most suitable for this comparison precisely because of it.

[91] Here, it once again needs to be noted that the research has not delved into the exploration of utterances that would qualify for said constructions, but which have for some reason – such as the one related to corpus-based nature of data extraction – been excluded from the analysis (e.g. relative constructions without an overt relativizer).

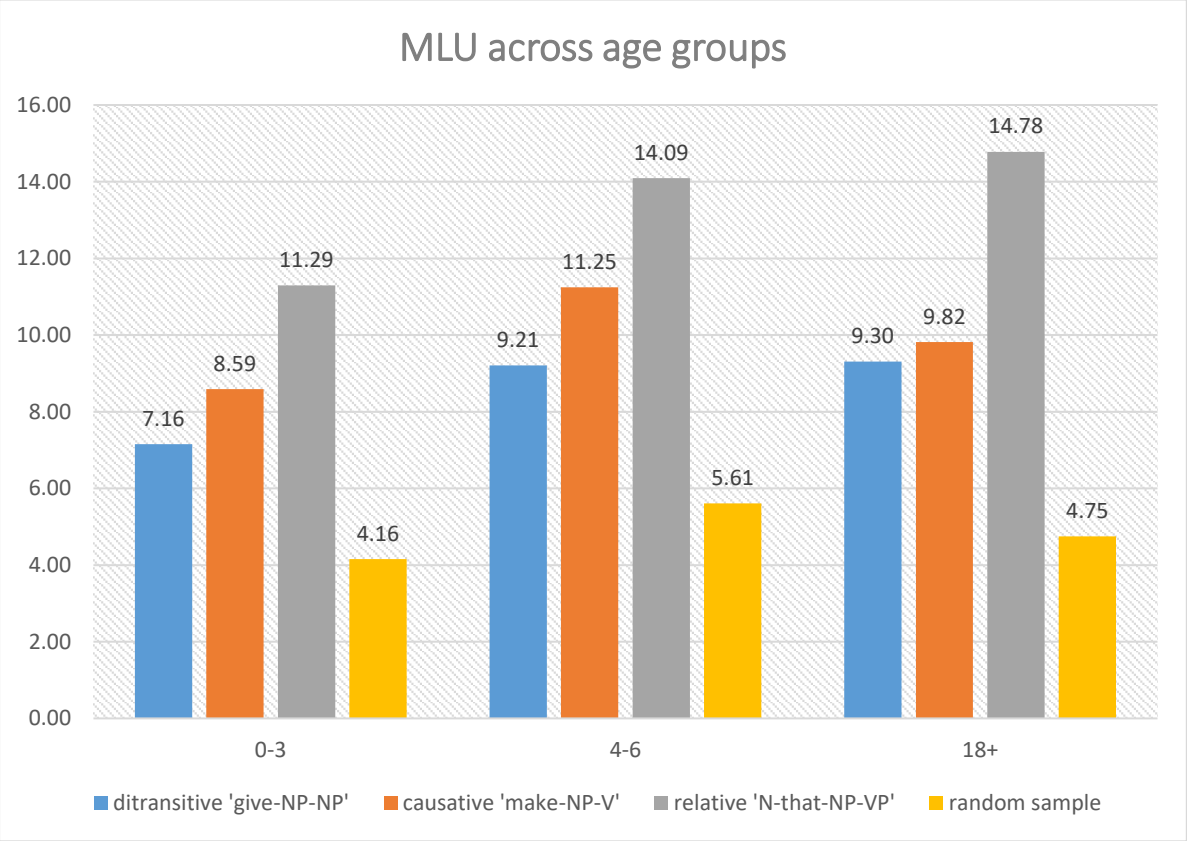
ditransitives and causatives in terms of overcoming the grammatical obstacles to producing them; although this study is not a longitudinal one, the sheer overall frequency of production indicates that these two types of constructions are almost equally challenging to the younger age groups. Obviously, there is no immediate link between the frequency of what has been said, and the difficulty of a construction in terms of processing. However, the fact that the comparable data suggests different trends of language production between the input and the youngest age groups is indicative of certain language acquisition aspects that hinder a more frequent production of those constructions which would otherwise be produced proportionately to their use in the immediate language environment. Since there is no particular pragmatic reason why ditransitives would be less, or equally, frequent in the younger age groups,[92] the discrepancy in terms of frequency seems best attributed to the processing difficulty of these constructions.

Additional confirmation that the processing difficulty partly reflects on the frequency of production is provided by the data on the length and lexical complexity. Firstly, the MLU measures exemplify the disproportion in the length of sentences between ditransitives, causatives and finally relatives (or at least, the patterns taken into consideration for calculating the measures of lexical richness and complexity). For the measures of lexical richness and complexity, it was first necessary to select some "representative" construction patterns for the comparative purposes for several reasons, the primary being the heterogenous lexical (and sometimes structural) nature of the observed constructions. The pattern give-NP-NP was selected as the most frequent ditransitive pattern. The pattern make-NP-V was selected as the most prototypical periphrastic causative pattern, and N-that-NP-VP was selected as a prototypical pattern of the more elaborate relative construction (i.e. apart from the relativized N slot, there is also the relative clause subject NP slot). The data calculated for lexical complexity indices in this section also includes a random sample of utterances that provides a clearer insight into the inter-constructional relations. The 'random sample' values were calculated based on a completely randomly selected set of sentences in CHILDES, which incorporates a wide range of linguistic patterns and elements produced by the observed groups.

[92] In fact, given that children tend to frequently express their need for something, ditransitive constructions should potentially be more represented than causatives; consider the expressions such as *give me food, tell me a story, make me lunch*. The amount of ditransitive constructions in total probably surpasses the overall use of periphrastic causatives, but the range of ditransitives in this thesis was limited solely to double-object constructions.

The rise in length is not equally proportionate across age groups, but the tendencies remain the same. The mean length of utterance increases both with age (the exception are the causative constructions between the 4-6 and the adult age group) and across constructions. The spike between causatives and relatives is more drastic than the one between ditransitives and causatives. First, the fact that these three differ according to MLUs both within and across different age groups, agrees with the initial assumption concerning the complexity of these constructions (causatives are more complex than ditransitives, and relatives are more complex than causatives). Though it has been claimed that the MLU is a gross indicator of language development, rather than an indicator of the structural complexity of grammatical competence, even when comparing the speech of children with similar rates of it (Jalilevand & Ebrahimipour, 2014), the claim is true for studies which examine a random collection of utterances with varying constructions at play, and this is precisely why a more detailed look into inter-constructional differences becomes a valuable addition to the discussion.

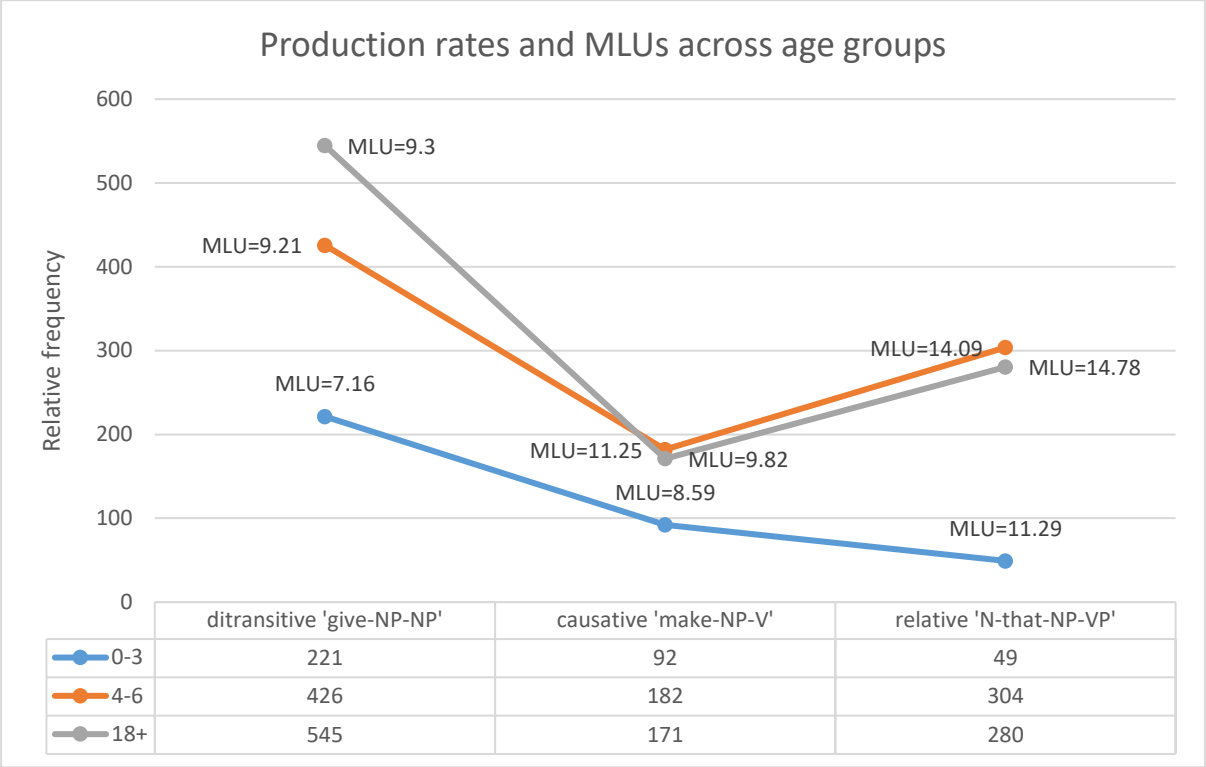
Figure 16. MLUs across constructions (morpheme-based)



Again, although there is no immediate link between construction complexity and length, there is evidence to suggest a correlation between length and processing difficulty (see Logan

& Conture, 1995). Second, the comparison of MLU data with frequency production data for the 0-3 and the adult age groups partly agrees with the claim ‘the longer the construction is, the less it will be produced’. This is especially interesting when one compares the production rates (relative frequencies) of the exact patterns on which the MLU was calculated. As already stated, it is important to be careful when comparing the production rates of specific construction patterns (the argument that some may be more or less produced for reasons completely unrelated to the syntax of the construction is even more compelling in the case of specific construction patterns), but the data on inter-age group differences in producing the three patterns on which the MLU was calculated additionally accentuates a peculiarity of the 0-3 age group (see Figure 17). The data in Figure 17 corroborates the claim that ties the MLU to the frequency of a construction by revealing a consistent drop in production rate of the analysed patterns with the rise of corresponding MLUs in the 0-3 age group regardless of the fact that this trend is not present in the 4-6 age group and the adult group. Even when it comes to the production rates of the observed constructions in general (not just specific construction patterns), the children in the 4-6 age group seem to produce almost equal amounts of the three constructions observed regardless of complexity or length differences (Figure 15). In fact, the same effect is probably not observed in the adult group since the nature of the speech is mostly child-directed, i.e. adults probably pay attention to the language they produce and avoid unnecessary complexity in talking to children. Nevertheless, The crucial insight gleaned from the data is that the production rates become less “burdened” with construction complexity after the child turns four.

Figure 17. Frequency of specific construction patterns across age groups and their corresponding MLUs

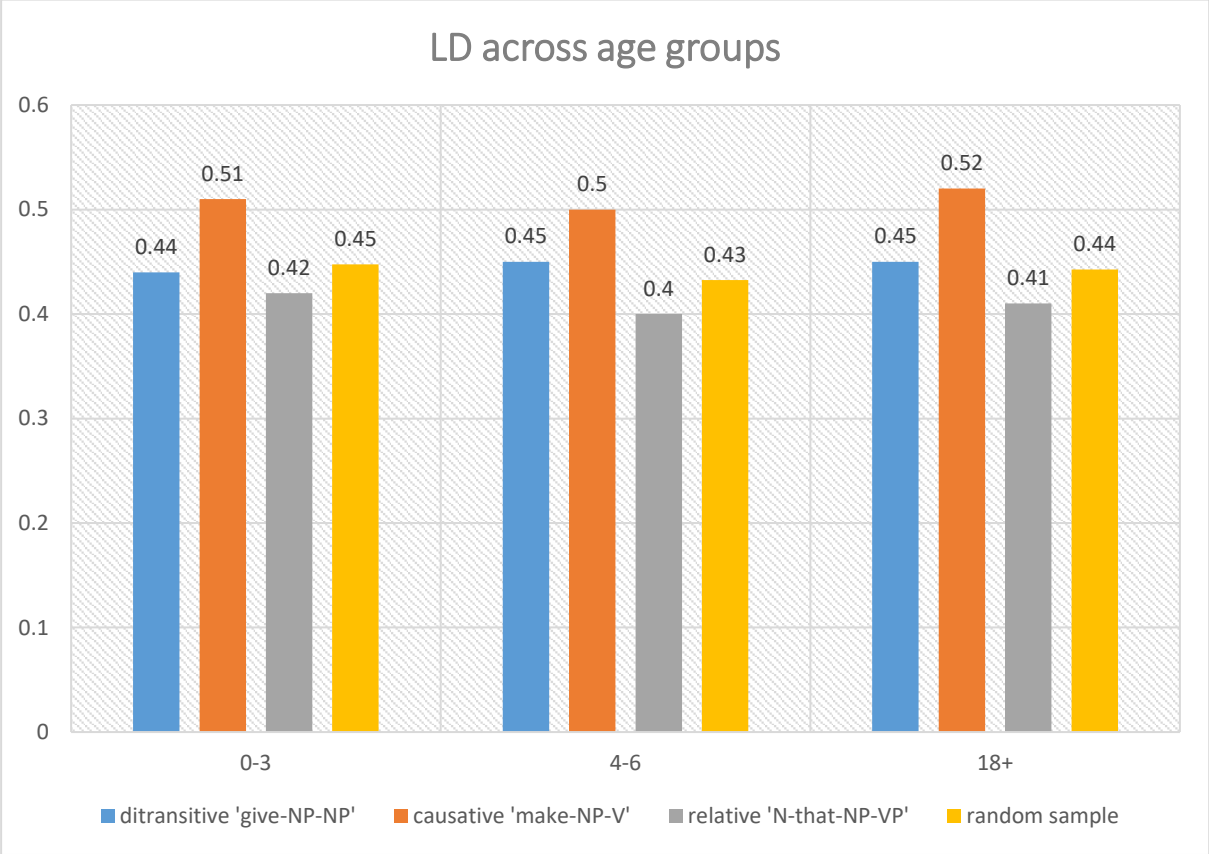


	ditransitive 'give-NP-NP'	causative 'make-NP-V'	relative 'N-that-NP-VP'
0-3	221	92	49
4-6	426	182	304
18+	545	171	280

When it comes to the basic measure of lexical density, there is little to no variation across age groups when it comes to utterances that contain the same constructions (Figure 18). In relation to each individual construction, the values are relatively similar for all three age groups. Instead, the differences are more evident between constructions themselves, rather than age groups, suggesting different levels of syntactic complexity. Within age groups, the drop is slightly more pronounced between the 4-6 age group and the 0-3 age group, which is in agreement with previous findings on the matter which attributed the low MLU values to the frequent use of monosyllabic words in the youngest age categories (see Li & Fang, 2011). Even when it comes to the random sample of utterances, the spoken language of all three age groups shares similar values, which is not necessarily unusual. The measure of lexical density, especially when it comes to impromptu spoken language, has been shown to vary very little across age groups (see Johansson, 2008). The greatest lexical density is characteristic for utterances centering around causative constructions, followed by ditransitives and finally relative constructions. The numbers are similar for ditransitive constructions and the random sample of sentences in each of the subcorpora, while relative constructions consistently score below them. As far as inter-constructional differences are concerned, the results are anticipated

to a certain degree; the relative constructions tend to employ more function words than both causatives and ditransitives, however, it is interesting that periphrastic causatives[93] exhibit greater lexical density scores than ditransitives on average considering that causatives allow elaborate sentences with several participants, all of which can take various types of determiners. In the same context, the LD scores for ditransitive constructions are probably lower due to more elaborate noun phrases (with more determiners).

Figure 18. Lexical density across constructions



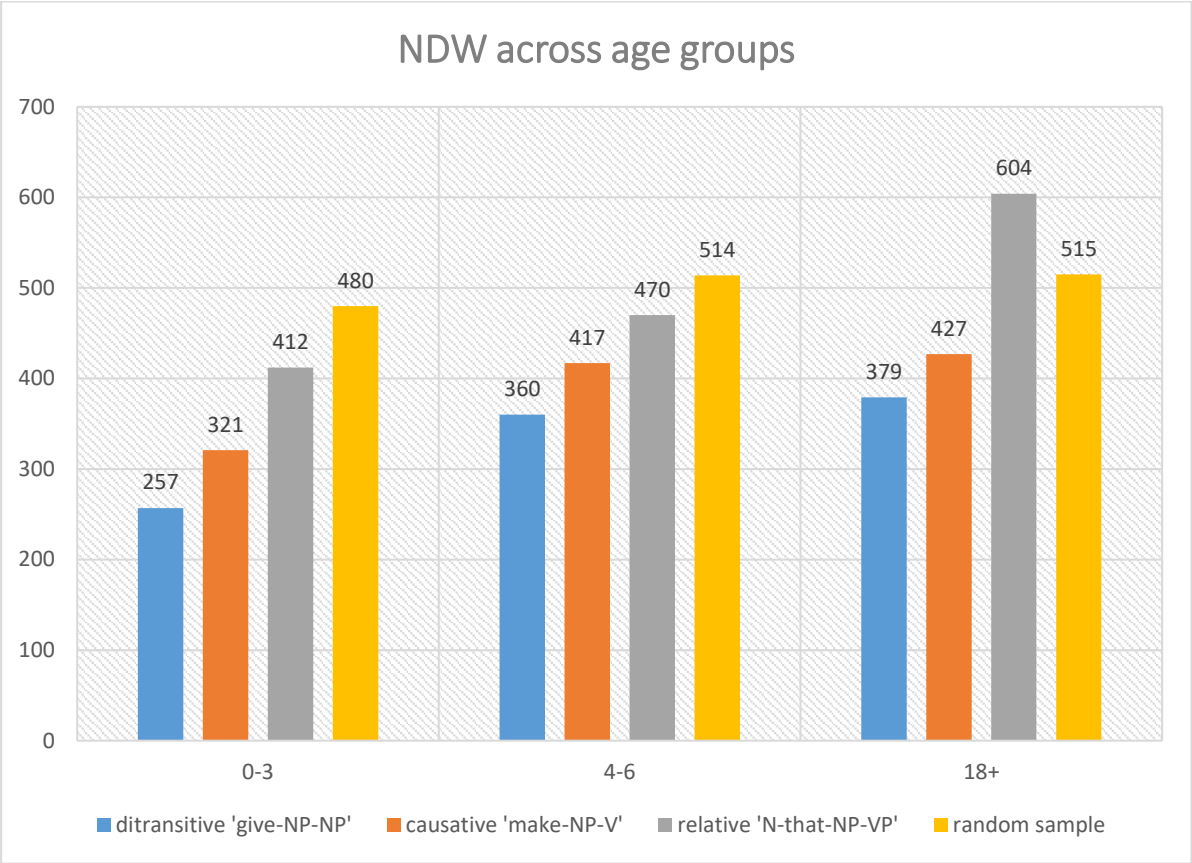
Johansson (2008) found developmental trends to be more noticeable for lexical diversity than for density, which, according to her, suggested "that lexical diversity is a better measure to use for detecting differences between age groups" (ibid., p. 77). One of the most reliable measures for lexical diversity is the number of different words employed in total (NDW; Figure

[93] At least those anchored in the verb 'make' where the caused activity is not expressed by the full infinitive. Had the calculation include the periphrastic causatives where the infinitival particle is taken into account, the average scores for the lexical density would be lower.

19). The data on this is particularly telling when observed across the three different constructions. Here, the differences are observable both across and within age groups. In the younger age groups, the greatest number of different words is found in the random sample of sentences on average, followed by relatives, causatives and ditransitives respectively. The assumed rise of syntactic complexity across the three constructions is accompanied by the rise of lexical diversity, while the high number of different lexical items in the random sample remains justified by the fact that these utterances are not anchored to any particular lexeme, or pattern for that matter. Ditransitive construction displayed the least number of different words per sentence even when the length of utterance is taken into consideration (see CCTR values in Figure 20). The results do not necessarily suggest that lexical diversity always rises as the construction becomes more syntactically elaborate, but the likelihood of the greater collocational character that is attributed to more basic expressions is confirmed at least in view of the constructions observed. In part, this relates to syntax-semantics interface of the constructions in question; for example, ditransitives constitute constructions whose syntax has clear inherent implications meaning-wise, whereas the relationship between meaning and grammatical pattern is less obvious in periphrastic causative constructions, and more rooted in the words themselves (such as the causative verb).[94] What is more, the ditransitive construction observed in this comparative analysis (*give*-NP-NP) should have relatively flexible object slots in terms of possible noun types that can occupy these positions; or at least, there is no reason to assume that the causative *make*-NP-V would be less lexically restricted than the observed ditransitive.

[94] In other words, the ditransitive meaning tends to be retained even when a nonce word replaces a head verb in a ditransitive pattern, thus indicating something inherently meaning-related about the argument structure of the ditransitive construction (cf. Goldberg, 1992), whereas the periphrastic causative construction is determined both lexically and structurally (the degree of causativity is strongly linked to the meaning of the causative verb, and can often be completely removed; cf. *he got him to move* and *he dared him to move*).

Figure 19. Number of different words across constructions

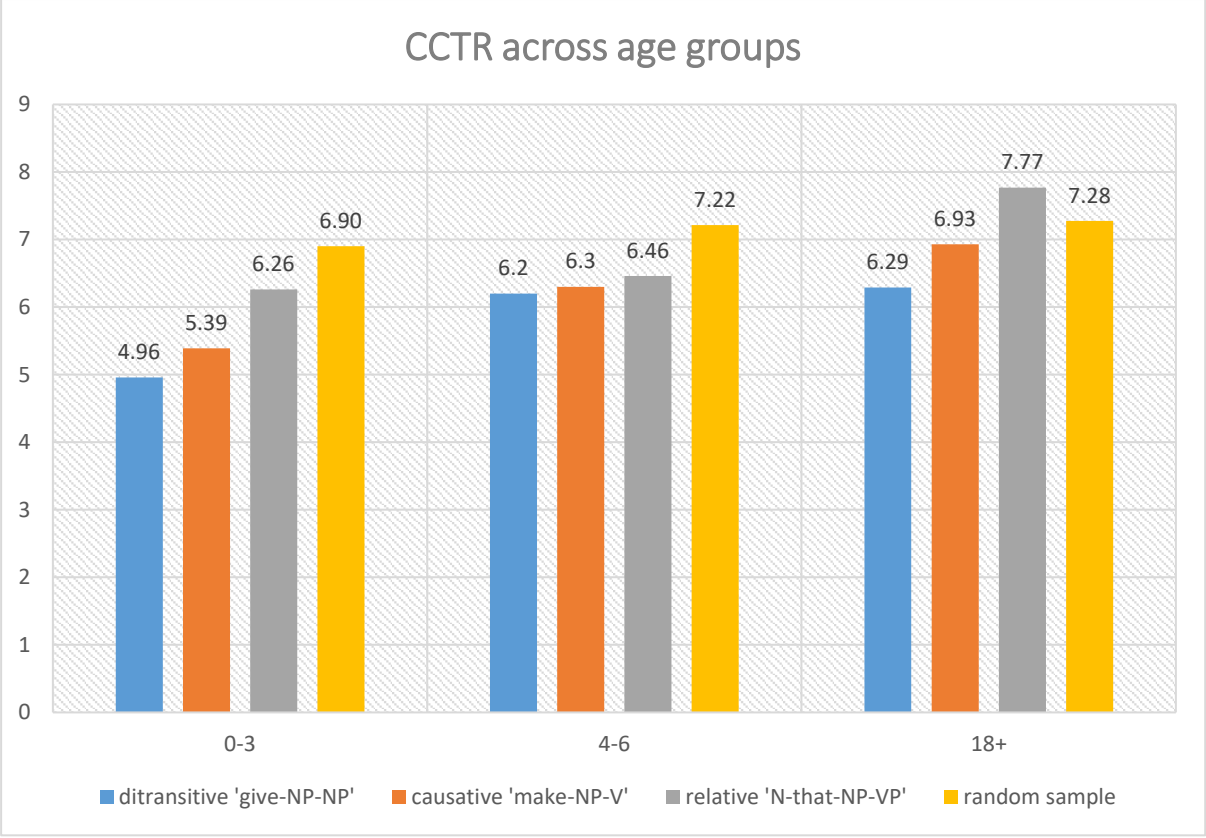


The most obvious discrepancy in the data are the relative constructions in the adult group, which seem to be the least restricted lexically, even going beyond the sample of random utterances. While the NDW values for the random sample were higher than in other utterances observed in the 0-3 and 4-6 age groups (480 and 514 respectively), the utterances with embedded relative constructions exhibited the highest NDW value in the adult group (604). This indicates a highly flexible use of relative sentences by the adult group, even in child directed speech. The increased level of syntactic complexity (compare LD values between relatives and the random sample) translates to an increased level in lexical diversity, but only in the adult group. In the two younger groups, even when the relative sentences are concerned, the numbers are higher for the random sample. In the adult group, the number of different words in the sample of relative sentences exceeds that of a random sample.[95] The results are in line with 'item-based' predictions about the character of early speech, even when it comes to

[95] This is partly caused by the length of utterance, which was shown to be the highest for the adult relative sentences in relation to other age groups. However, the rise in the number of different words is not proportionate to the rise in sentence MLU; it is clear that adults are indeed more lexically diverse in their use of relative sentences, even in child-directed speech. The issue is further addressed with type-token ratios for each group.

constructions such as relatives, where the construction slots are open for various types of nouns and verbs. This means that, as soon as the analysis in the younger age groups restricts itself to a certain grammatical pattern, the NDW measures reveal lexically more restricted language. The finding is interesting because it may suggest that the item-based character is not necessarily confined solely to verb-restricted constructions and those lower in schematicity (those lower in the number of different types of arguments permitted), but that it appears even in patterns such as relatives (which are syntactically more complex and arguably more flexible with regards to lexical items in key construction slots).

Figure 20. CTTR across constructions



As far as lexical diversity is concerned, CCTR values further confirm the suppositions concerning the range of words used and the interpretation thereof. Since type-token ratio takes into account the total number of tokens in the utterance (in other words, the size of the sample),[96]

[96] Though utterances were selected randomly for comparison, and though their number was equal across constructions (see Section 2.1.3), there were some obvious disproportions in utterance length (the number of tokens

the results further justify the interpretation of increase in lexical diversity as a side effect of the increase in syntactic complexity. The type-token distribution (CCTR; Figure 20) corresponds to that of NDW (Figure 19), although the values for the 4-6 age group are slightly more balanced out across the three constructions.

However, the most intriguing aspect of both distributions is the similarity in values between the random samples across all three age groups (CCTR was 6.9 in the 0-3 age group, 7.22 in 4-6 age group, and 7.28 in the adult group), and the corresponding dissimilarity in other constructions. This data is especially relevant in the context of language acquisition and development as it strengthens the argument that early grammatical patterns centre around less items than they otherwise would in adult speech. In other words, the commensurate values for the random sample of sentences indicates close-to equal lexical diversity exhibited by age groups (though still slightly skewed in favour of adults), which in turn should be emulated in the use of particular constructions as well; however, it happens that the values of lexical diversity for constructions are considerably lower in the younger age groups – especially in the 0-3 age group, which is the one predicted to be the most prone to the item-based character of early construction use.

A large portion of data in this thesis concerns the proportions of particular syntactic patterns and lexical items in the observed positions. This has been deemed relevant in relation to both input-output correspondence and the degree of item-based character in the produced constructions. Although the data clearly indicate the degrees to which the specific syntactic patterns within constructions are produced, it remains unclear to what extent the dominant syntactic category in the 0-3 age groups differs in its prominence when compared to other age groups. For example, the saturation of the most frequent relative pattern is the highest in the 4-6 age group, which outscores the 0-3 age group by several percentage points, while the most frequent ditransitive pattern exhibits a more distinctly dominant rate in the adult group (in relation to other syntactic patterns produced within the same age group). While there is no doubt that the production rates of certain constructions coincide remarkably across all age groups, it is not always the case that the dominant syntactic patterns have higher production rates in the youngest age group. On the other hand, the item-based tendencies which concern the lexical

per utterances); i.e. the more elaborate the embedded construction, the longer the utterance containing it tended to be.

character of the constructions produced are better explored in the context of the proportions to which the most frequent items are produced. When the data is systematized, it is easier to observe whether there is a more pronounced tendency of the youngest age groups to overproduce certain lexical items in the observed constructions slots (see Table 39).

Table 39. Overview of the most frequent items in each construction[97]

Construction	Construction slot	Age group	Dominant lexical item	%	Dominant category of animacy	%
ditransitive *give*-NP-NP	1st (indirect) object	0-3	*me*	52%	Inherent A	79%
		4-6	*me*	49%	Inherent A	83%
		18+	*you*	29%	Inherent A	70%
ditransitive *give*-NP-NP	2nd (direct) object	0-3	*that* (DEM)	8%	Inanimates	89%
		4-6	*some*	5%	Inanimates	94%
		18+	*kiss*	11%	Inanimates	94%
causative V-NP-V	object argument (causee)	0-3	proper names	18%	Inherent A	44%
		4-6	proper names	11%	Inanimates	44%
		18+	mommy variants	13%	Inanimates	47%
relative N-*that*-NP-VP	relativized noun	0-3	*thing(s)*	15%	Inanimates	83%
		4-6	*thing(s)*	21%	Inanimates	81%
		18+	*thing(s)*	9%	Inanimates	83%
All ditransitives	ditransitive verb	0-3	*give*	43%	n.a.	
		4-6	*give*	42%	n.a	
		18+	*give*	37%	n.a.	
All causatives	causative verb	0-3	*let*	67%	n.a	
		4-6	*let*	64%	n.a.	
		18+	*let*	59%	n.a	
All relatives	relativizer	0-3	*that*	67%	n.a.	
		4-6	*that*	74%	n.a	
		18+	*that*	68%	n.a	

Again, the answer to the question of whether the most frequent categories in adult speech are also more prominent in the 0-3 age group is not completely transparent, that is, the majority of the most frequent items observed in this analysis across age groups are also the ones

[97] For lexical items, the proportion of use is calculated on the basis of their frequency divided by the total number of tokens, whereas for animacy the proportion of use is type-based (type frequency of animate or inanimate stems divided by the total number of types). For a token-based calculation of animacy, see Appendix 10.2.

that have the highest rate of production within 0-3 age group, with the exception of the relativized noun and the relativizer, which are more saturated within the 4-6 age group (although, the relativizer has only few items to compete with when compared to the rest). The item-based phenomenon for verbs heading the constructions has already been observed and hypothesized by other researchers (Tomasello, 2000b, 2001c; MacWhinney, 2005, 2014; Hodges et al., 2004), and the item-based character is confirmed in these corpus data as well – at least when it comes to the production rates of ditransitive and causative verbs. It also seems that, even in relation to child-directed speech, the item-based tendencies can extend to nouns as well in the youngest age groups, but data is not consistent for all construction slots. As far as type-based animacy is concerned, the data does not confirm the prediction that the most dominant category will also be the one that is most distinctly overproduced in the 0-3 age group. This goes against the hypothesis that children would seldom go beyond the "prototypical" semantic features of the items (not just the exact items they heard) in their (re)production of the heard construction slots. To be fair, the degree of overlap in animacy is as impressive as with some of the lexical items, and it would be wrong to completely reject the supposition that the children's reproduction of the input in some sense mirrors the properties of items that children were able to deduce from their immediate linguistic environment. In this context, it is particularly interesting to observe the causee construction slot, which takes animate items as their most frequent exemplars in the input (*mommy* and *daddy*), but which does not reflect on the overall animacy calculated on the basis of different types occurring in the slot. Instead, the most frequent type-based category of animacy is occupied by inanimates for both adult speech and 4-6 age group, but in the youngest age group, the category is most dominantly occupied by inherent animates. This might be an indicator that the children's deduction of the properties of items that can occupy the target construction slots is primarily reliant on the most frequent exemplars heard in said slots. Moreover, this also represents the construction slot where the adults have employed the greatest range of items with respect to animacy, and it is also the one where the children were not as restricted with their use of animate types as with other constructions; i.e. they were not employing one type of item as predominantly as with other construction slots. The overall takeaway seems to be that the children are indeed flexible in their production exactly where the adults are flexible in their production, but there does not appear to be any intra-group "anomalies" in the 0-3 age group in comparison to others.

As far as lexical similarity (and dissimilarity) is concerned, the differences between the 0-3 and the 4-6 age groups across the three constructions also revealed some aspects of cognitive, and not just linguistic development. For instance, when it comes to the slightly more abstract *show* and *tell* in ditransitive construction, *tell* in the periphrastic causative construction, or the use of the relativizer *when* in the relative construction, it is obvious that the tendencies observed within constructions across age groups deviate when it comes to these items. The fact that they seem to be underrepresented (or represented less than expected considering the frequencies of other items observed in the same constructions) in the 0-3 age group, and appear to "get back on track" in the 4-6 age group might further showcase that cognitive development is a prerequisite for the linguistic one. The verbs in question tend to be conceptually more abstract (in themselves) than others in the same constructions, and the relativizer *when* associates with the temporal dimension which is again more conceptually abstract than other relations in the relative constructions. The data fits into the interpretation that the ability to grasp more abstract concepts and to apply this knowledge linguistically marks an important step in the cognitive development of a child – a step which becomes most evident after the children turn 4 years of age (cf. Wellman et al., 1995; Tomasello, 2005; Clark, 2009, p. 169).

Furthermore, the data on the proportions of different lexical items that have been observed in the input (the adult group) is very telling in terms of which age group corresponds more in the choice of said items in the specified slots (see Figure 21).[98] It is clear that, regardless of the construction observed, the part of speech, or the degree of syntactic relevance that a specific slot within the construction bears, the data indicates a trend where the 0-3 age group uses items heard in the input, in the same slots where they have heard it being used, to a greater extent than the 4-6 age group. In other words, the proportion of lexical items that were also observed in the input was greater within the 0-3 age group than those in the 4-6 age group. Furthermore, it is also interesting to observe the inter-constructional differences with regards to the degree of item correspondence in the specified slots.[99] It is immediately clear that the "simplest" construction is also the one where the children/adult overlap is the highest; more specifically, more than 80% of the items employed in the position of the direct object by the 0-

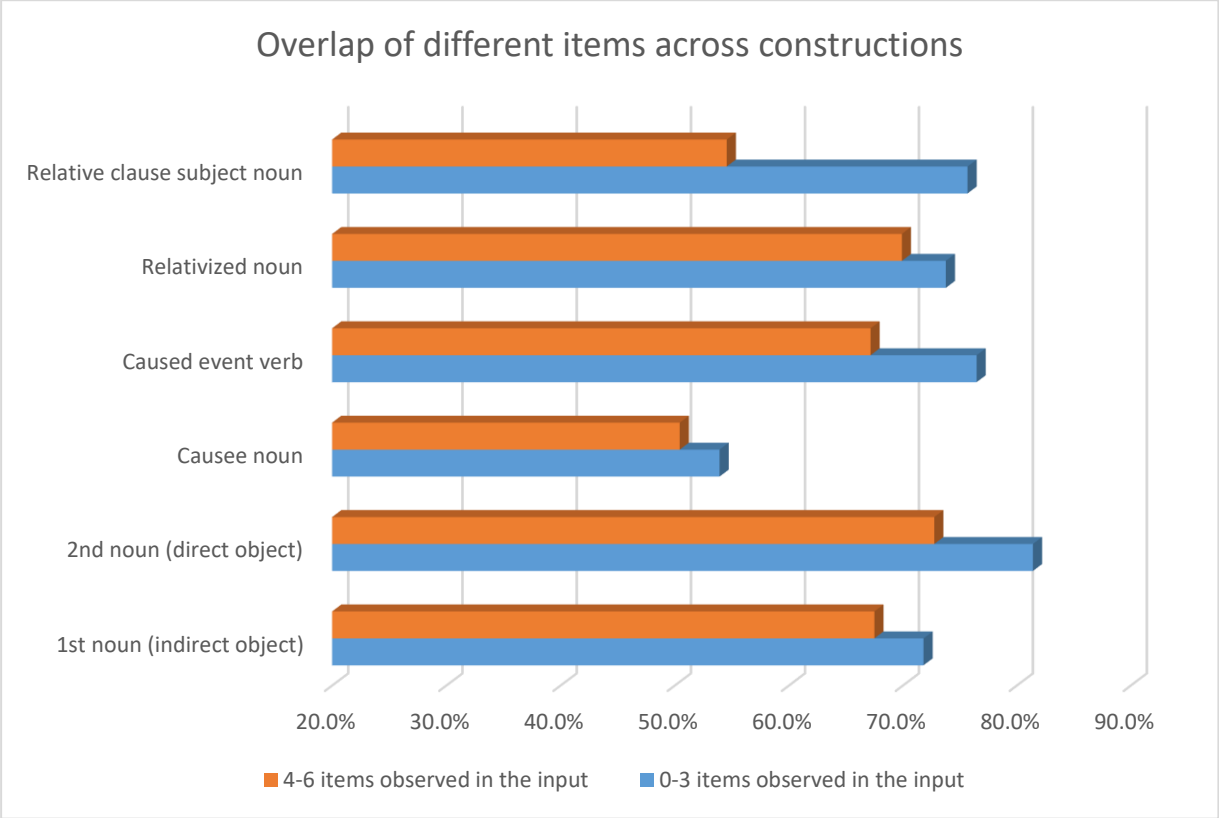
[98] This is an additional benefit of corpus research, as it allows the comparison of particular patterns on a larger scale with a time-wise relatively low cost in comparison to experimental designs.

[99] The comparison of specific syntactic slots in-between constructions is problematic for a number of reasons, most of which relate to specific construction constraints and the diverging ranges of items that usually occur in those instances, but a more 'across-the-board' approach sheds light on the relative extent to which children rely on their immediate language surrounding.

3 age group were also observed in adults, whereas other construction slots do not exhibit such degree of overlap (though they are closely followed by caused event verb slot and the relative clause subject noun slot). In relation to the other two constructions, ditransitives are acquired earlier in development, meaning that we might expect a smaller discrepancy between age groups in the noun slots (mastering the construction earlier also entails that one would become more diverse with it earlier in development), but this is not completely evident; we find equally small discrepancies in the causee noun slots or relativized noun slots.

In terms of other items and their input/output overlap, two categories stand out in particular: the causee slot and the relative clause subject slot. The overlap in the causee slot is relatively surprising when compared to the rest, as it does not necessarily imply a greater degree of flexibility in the use of items that can occupy the slot; in fact, it seems to carry more constraints than some other ones as it prefers animates as opposed to inanimates (for example, this is not the case with the relativized noun slot, which nevertheless exhibits a greater degree of overlap in use). On the other hand, the relative clause subject category is intriguing because it displays the greatest discrepancy between the 0-3 and the 4-6 age categories compared to other construction slots. This suggests that the children's production of object relative clauses is deeply tied to the input in terms of certain lexical choices, but also that there is a significant move in diversity as they turn 4.

Figure 21. The proportions of different lexical items in specific construction slots that have been observed in the adult group (input)



Most importantly, the results confirm the hypothesis stating that the 0-3 group will resemble the language of the adults more than that of the 4-6 age group. The data reflects the idea that input is more crucial in the earliest stages of development in terms of 'imitative learning'. This is not to say that general cognitive mechanisms do not play a role from the start, or that 'imitative learning' is simply a reproduction of what has been heard. Instead, the data highlights the effect to which the role of the immediate language enviroment bears in early language production, and suggests that other cognitive skills responsible for the development of a more comprehensive language proficiency come into play later in life, allowing the speaker to use a wider range of items in the particular slots. In other words, as they grow older, their production starts to go beyond the previously experienced exemplars in the target slots, employing items whose linguistic character appears suitable to fill the slot – the suitability inferred by children that is underpinned by domain general processes.[100]

[100] Association, recognition, memory retrieval and generalisation underly not only one's cognitive development or language acquisition, but also the development of more complex cognitive skills. More importantly, they have

Another important observation that has revealed itself across all three constructions observed is the fact that there has always been one lexical element[101] that dominated the production of the studied constructions. There has always been an element common for the majority of the produced linguistic patterns. In ditransitives, this is the verb *give*, but it is also the pattern with the pronoun in second position and the complex nominal in third. In causatives, the clear winner is the 'let construction', while the majority of relative constructions rely on *that* as the main conjunction between the two clauses. What is more, the resemblance between age groups was not exclusive to lexical similarity, but also to the syntactic one. Although these two have been intertwined in some parts of the analysis (part of the reason being that the combination of syntactic arguments and lexical items needed to be exploited for the extraction of data at times), there is no doubt that patterns which tended to be most frequent in terms of both lexis and syntax in one age group, also tended to be most frequent in other groups as well. The question is what can this item-based tendency tell us about these constructions in general in the context of language development? Although the direct link between a particular construction's volume of production and the facilitation of development is beyond the scope of this work, the fact that the phenomenon of lexical similarity tends to constantly reiterate itself in different constructions and contexts goes to show that the input-output relationship is by no means coincidental. While it is not surprising that children reproduce what they hear, the results shed some light on the degree to which this 'imitative learning' occurs, especially in relation to different constructions and language patterns.

6.2. Final discussion

Chi Square values always indicated that the age groups significantly differed in terms of the overall production rates of different construction patterns,[102] while the Spearman's rank correlation coefficients confirmed a very high degree of similarity in terms of frequency

been argued to account for the fast-mapping that is central in children's lexical learning (cf. Saffran & Thiessen, 2007).

[101] Here, the term also encapsulates some elements typically regarded as grammatical (or functional) ones, such as relativizers (or the relativizer *that* in this specific case), which, due to the (often) meaning-loaded nature of their role in the relative construction (e.g. *who* can only modify sentient beings and *which* non-sentient objects) overlap with proper lexical items in some respects.

[102] The analyses were conducted on normalised/relative frequencies, meaning that the difference in frequency reflected the proportion that the observed constructions occupy in the totality of their recorded speech.

hierarchy; i.e. in terms of overproduced to underproduced structures, the intra-group preferences were nearly identical frequency-wise across age groups. The former results can mistakenly be taken for granted as the result of the fact that adults are more linguistically capable or syntactically rich in their expression, but it is not obvious that these would necessarily translate to the fact that certain structures are more frequent in the speech of those who are more linguistically proficient, and this is especially true for simpler patterns. That is, the production frequency rates should reflect the syntactic simplicity of the observed construction; if anything (considering that the frequency is normalised) simple structures should be more frequent in the youngest age groups because they would make up the majority of propositions. In that sense, the data actually comes as a confirmation of the counterintuitive take on corpus frequency as a good indicator of linguistic proficiency; for example, in relation to the other two "simpler" constructions, 0-3 age group most drastically differs in the relative construction production rates, exhibiting the lowest numbers.

As stated, equally relevant were the results pertaining to the Spearman's rank correlation coefficients, which showed, when the observed patterns were hierarchically arranged, that the inter-group correspondence was remarkably high, indicating that the patterns overproduced in the input would also be overproduced in the output (and the same goes for the underproduced ones). However, the correspondence was high across all age groups, meaning that the major claim of the first hypothesis – stating that the 0-3 age groups would be more similar to child-directed adult speech than the language of age groups 4-6 when it comes to the ratio of the use of certain structures – is not fully confirmed; the correlation was too high among all three age groups in terms of intra-group production hierarchy trends of the observed construction sub-patterns. On the other hand, when it comes to lexical correspondence, the results have been much clearer. The overlap in the observed construction slots has been rather one-sided, regardless of the inter-constructional discrepancy in complexity. The results convincingly indicated that the speakers of the 0-3 age group are more conservative in their language use than the 4-6 age group, that is, a greater proportion of items used is observed in the input across all of the construction slots observed (whether nouns or verbs) in the 0-3 age group, which (as predicted) signals that the group sticks more firmly to the input in their (re)production of language, which seems to be in line with some of the previous research on the matter (see Liu et al., 2008; Goodman et al., 2008; Shin, 2022).

When looking at the data in total, across all age groups, especially those data concerning the number and ratios of particular patterns produced, but also that concerning particular items (especially verbs) and noun animacy, the language acquisition counterpart to the 'Pareto Principle' comes to mind. Across all age groups, the majority of produced utterances are saturated around one or two patterns, which – despite competing with a number of patterns that could be alternatively used for similar/same conversational or expressive purposes – typically cover at least 80% of the overall use. Moreover, the concentration of the produced patterns and items is always, albeit only slightly, more accentuated in the younger age groups. The corpus data concerning the similarity of certain lexical items in specific construction slots confirmed the first hypothesis predicting a considerable lexical and structural similarity in constructions across age groups in the observed constructions. Moreover, it confirmed that the language of the 0-3 age group is more conservative than that of the 4-6 age group, exhibiting greater use ratios of items observed in the input in the same slots. Not only does this agree well with the 'item-based' approaches to language acquisition, but it also agrees well with Ambridge and Lieven's (2015; pp. 494-495) description of the trajectory that language acquisition takes in terms of early production. Children make mistakes when learning language because different forms of language compete with each other, from individual words to abstract expressions. Initially, children might use the wrong form of language because it may be more common or they have not learned the appropriate alternative yet, but as they learn new ways to express themselves that fit their intended message better, they move away from making mistakes. This learning is thus impacted by statistical tracking of instantiations of interest and semantic mapping of construction slots. Children learn the probability of certain word combinations and constructions, which causes some constructions to be used more often. Additionally, they learn the specific meaning of different construction slots, which makes some items more suitable for certain constructions than others. Therefore, as the frequency of these items in non-standard constructions increases and their semantic compatibility with the target construction improves, it becomes less likely for children to make errors and more likely to understand the intended meaning.

The second hypothesis – predicting that 'item-based' tendencies will translate to the semantic features of the items produced by the youngest age group – is not confirmed, i.e. while there has been a striking similarity in terms of the degree to which the young age groups correspond to adults even when it comes to the type of items produced, there has not been a

clear indication that the most frequent category of animacy in the input is also the one that exhibits more dominant production rates in the 0-3 age group than in 4-6 age group. The similarity in frequency rankings of specific lexical items or construction patterns fits well with the usage-based claims about the nature of language acquisition, but according to the item-based claims, the first few most frequent lexical items or construction patters would take up an even greater portion of the overall production in the youngest (0-3) age group. If we assume that the same logic can be applied to the semantic features of lexical items (not just lexical items themselves or construction patterns), the item-based predictions would assume a more conservative use on behalf of the 0-3 age group in terms of some semantic properties – such as animacy of the items occupying specific construction slots (after all, it is possible to expect that the item-based tendency in the use of specific lexical items would translate into similar trends when it comes to their semantic properties). More precisely, the production trends of specific animacy values in specific construction slots correspond across age groups, but without a clear indication that the most dominant animacy values in general are even more dominant in the 0-3 age group when compared to others). The hypothesis was anchored in the supposition that children's representations of specific language patterns closely tie to the semantic features of the same constructions which they encounter most often in the input (Buckle et al., 2017), but the data clearly indicates similar intra-group and intra-construction tendencies with respect to animacy, and yet low inter-group deviation in view of over/under production of specific categories. The data does indicate that all three age groups tend to over- and under-produce items with corresponding animacy values to similar degree, which does not support the expectation that the youngest age group would solely stick with the prototypical animacy value for a particular construction slot. It is not that they do not predominantly produce what has also been mostly heard in the input, but the predominance of one category within the 0-3 age group should be even more apparent than in the 4-6 or the adult age group. This however, does not disprove the fact that the children's learning of a construction is also distributional in character, especially considering that the construction slots – where the input exhibits the greatest range of items with respect to lexical properties (such as animacy) – are also the ones where the children do not employ one type of item as predominantly as with other construction slots. According to data, there is little doubt that children deduce the "appropriateness" of lexical properties that are to occupy certain positions within a construction by taking into account all of the various exemplars and instances which they have heard them in, but the corpus data also

does not indicate anything particularly specific about the distribution pattern of animate items in the 0-3 age group in comparison to the other two.

The third hypothesis was that the measures of lexical richness and complexity would differ across age groups despite the potential lexical similarity anchored in the 'imitative' and 'item-based' character of language development. The first thing to note is that the measures of morphological elaborateness and lexical diversity reflect the inter-constructional differences in terms of the initially assumed complexity gap (particularly useful in this regard are the values of the random sample of utterances). The values pertaining to the MLU, type-token ratios and the number of different lexical items employed, demonstrate an ascending inter-constructional trajectory – relative constructions are shown to have higher values than causative constructions, and causative constructions are shown to have higher values than ditransitive constructions. More importantly, the results show not only inter-constructional differences, but inter-group differences as well, with the indices consistently increasing as the age progress. Arguably, inter-group differences are less pronounced than inter-constructional differences, and the gravity of the extent to which they vary is subject to discussion. It is certainly obvious that the inter-group jump is always more substantial between the 0-3 age group and the 4-6 group, than between 4-6 and the adult group (where the difference is either small or absent; e.g. see MLU values). In part, these results agree with those of Blake, Quartaro and Onorati (1993, p. 150), who found MLU to be a much more reliable measure of clausal construction complexity in the younger age groups (from single-word utterances, two-constituent clauses to clausal connectivity), but with a declining applicability in the more advanced stages of children's language development. Moreover, the intra-constructional data aligns well with the suppositions of MLU values being lower in the younger age groups due to the frequent use of monosyllabic words in the youngest age categories (see Li & Fang, 2011). The absence of more noticeable differences between 4-6 age group and adults has mostly been attributed to the fact that the adult speech in CHILDES is mostly child-directed, and this reflects most on the values in more complex constructions (the argument being that adults remain aware of the complexity of those constructions and that this is where they moderate their speech the most; e.g. they moderate it more with relatives than they do with ditransitives, hence the lower difference in values). Overall, the hypothesis is confirmed, and the measures of both complexity and diversity are relatively reliable indicators of language development, though this is more obvious in the comparison of the 0-3 and the 4-6 age groups.

The final hypothesis, which predicts that simpler constructions should exhibit a greater level of item-based character than more complex constructions, ties well to the previous hypothesis but also requires a multi-layered answer. One of the major claims of the usage-based account is that children's early productions of utterances revolve around a restricted number of items, that is, concrete and specific items or phrases, rather than structural categories or schema-based competence, and that abstract and adult like categories emerge in a gradual fashion later on (see Tomasello, 200b, p. 156). If this is correct, a corollary of this claim would be that more complex linguistic constructions should also be less item-based, which is not obvious from the data obtained in this study. The first and the main indication of the decreasing item-based trend in more complex constructions are the measures of complexity and diversity, which confirmed both inter-constructional and inter-group differences even in terms of type-token ratios which negate the effect of utterance length (otherwise, the increase in the range of lexical items could be attributed to the fact that more complex constructions tend to be longer). However, the claim of 'item-basedness' was also put to test in relation to the concentration levels of the most frequent items that occupied the observed construction slots, and the input-output overlap in them. This data was more telling of the fact that all constructions across all age groups tend to be item-based to a certain degree. In fact, there were striking similarities when it comes to specific lexical items being employed in the corresponding construction slots even when there is plenty of competing lexical items that could otherwise take their place, but it is difficult to deduce inter-constructional differences with respect to how concentrated the most frequent categories are (whether when it comes to some exact lexical items or categories of animacy). The data on the overlap between the younger and the adult groups is more suggestive of the 'imitative' character than it was of 'item-based' character of early speech; with the exception of the causee noun slot, the degree of overlap was similarly high in all of the observed construction slots despite their syntactic elaborateness. In total, the data does not seem to suggest a greater degree of item-based character in the simpler constructions, but instead exposes it as a trait of all constructions.

Naturally, there are several obvious limitations to the data that has been derived this way which significantly affect the interpretation of the results and reduce the conclusive potential of the findings. First, the obvious limitation of this type of corpus research is that it observes only the language that has been produced (i.e. speech), and the gap between language production and language proficiency arguably translates to the gap between language

production and language acquisition. The second obvious limitation to the study is that we only possess a segment of the children's language, and, even more importantly, what has been referred to as the 'input' in this study, here represents only a small portion of what the children have been exposed to in their lives. In this context, it is often argued that studying differences or negative linguistic evidence may be more advantageous than studying similarities in language acquisition. However, in the context of the CHILDES corpora, this approach could lead to more methodological challenges rather than solving them. As a result of the lack of complete knowledge of the input the child has been exposed to throughout their life, it becomes methodologically disingenuous to interpret the child's non-canonical formulations and seemingly creative utterances as deviations from what they have heard, consequently inviting questions of what constitutes a sufficiently comprehensive dataset for deriving such conclusions. Finally, some might suggest that the drawback of the study is the heterogenous nature of the data and different ways in which the children's speech was elicited by the researchers, but this is not necessarily as damaging to the findings as it might be suggested. Many of the lexical choices are doubtlessly a product of the circumstances in which the interactions were recorded, but the fact that the speech reflects certain situational or environmental aspects of said interactions, or even behavioural or cognitive aspects of participants given the age differences (such as differences in self-centeredness, awareness and self-awareness), is not to be treated as a drawback to data, but as its main advantage. While there are considerable limits to the conclusions that can be drawn solely from linguistic data, the fact that it reflects various factors that play a role in the complex phenomena of language production and early development is valuable in itself. To interpret the corpus-derived data often means going beyond the immediate product that is being observed. Even experimental research designs show that it is difficult to isolate certain effects that mediate the process of language processing, and it seems impossible to study language development and acquisition in isolation of situation-specific and participant-specific aspects that affect both language and linguistic data. Instead, the findings such as the one presented in this research should serve to complement a number of existing experiments and studies in the field of language acquisition, and contribute to the totality of data that seems best interpreted as a whole, rather than in isolation from one another.

7. CONCLUSION

In some sense, the comparisons obtained in this study can be reduced to four different types: intra-constructional, inter-constructional, intra-age group and inter-age group observations. Inter-age group and inter-constructional observations are more relevant for addressing the research hypotheses, the first of which predicted lexical similarity on several levels and implied a more pronounced 'item-based' tendency in the youngest age group. In general, the data across all constructions tends to exhibit what can be described as a linguistic version of the Pareto Principle, where most utterances primarily rely on a few common language patterns and items. The first hypothesis has generally been confirmed across several levels, indicating a considerable similarity in lexical items and construction patterns across age groups, suggesting a more conservative language use and more pronounced concentration of specific items and syntactic patterns with the decrease of age.

The second hypothesis, which predicted that the youngest age group would predominantly produce items with a prototypical value of animacy in observed constructions, was not confirmed. There was a noticeable similarity in the type of items produced by the younger age groups compared to adults, but it did not clearly indicate that the most frequent category of animacy in the input was more pronounced in the 0-3 age group. This lack of inter-group dissimilarity also meant that intra-group production rates were similar across age groups. All three age groups tended to over- and under-produce items with corresponding animacy values to a similar extent, contrary to the expectation that the youngest age group would stick with the prototypical animacy value. Naturally, this does not go against the fact that children's learning of construction is distributional, as they seem to consider various exemplars and instances from their input to deduce the appropriateness of lexical properties in certain positions within a construction, but the degree to which they do it does not apparently affect the distribution of said items with respect to animacy. In other words, the data has not clearly indicated a specific distribution pattern of animate items in the 0-3 age group compared to the other age groups.

The hypothesis predicting differences in lexical richness and complexity across age groups, despite potential lexical similarity in language development, was confirmed. Likewise, the measures of morphological elaborateness and lexical diversity revealed inter-constructional

differences, with relative constructions having higher values than causative constructions, and causative constructions having higher values than ditransitive constructions. Additionally, the inter-group differences were also rather detectable, with the indices of lexical and syntactic elaborateness consistently increasing as the age progressed. The most substantial jump in inter-group differences was observed between the 0-3 age group and the 4-6 age group, while the differences between the 4-6 age group and the adult group were either small or absent. This was attributed to the fact that adult speech in CHILDES is mostly child-directed, particularly in more complex constructions, where adults moderate their speech, resulting in lower differences in values. In total, the measures of complexity and diversity were reliable indicators of language development, particularly evident when comparing the 0-3 and 4-6 age groups.

The hypothesis that simpler constructions would be more item-based than more complex constructions was not strongly supported. The analysis of complexity and diversity measures indicated differences both between constructions and age groups, even when accounting for utterance length. However, the examination of specific lexical items used in construction slots showed that item-based tendencies were present across all constructions and age groups. Specific lexical items were consistently employed in corresponding construction slots, even when alternatives were available. Additionally, the data on overlap suggested that early speech leaned more towards an 'imitative' character rather than being solely item-based. The degree of overlap was high across all construction slots, regardless of their syntactic complexity. In summary, the data did not clearly show a greater degree of item-based character in simpler constructions; instead, it indicated that this trait was present in all constructions to a relatively similar degree.

8. BIBLIOGRAPHY

8.1. References:

Ai, H., & Lu, X. (2010). A web-based system for automatic measurement of lexical complexity. In *27th Annual Symposium of the Computer-Assisted Language Consortium (CALICO-10). Amherst, MA. June* (pp. 8-12).

Ambridge, B., Bidgood, A., Twomey, K. E., Pine, J. M., Rowland, C. F., & Freudenthal, D. (2015a). Preemption versus entrenchment: Towards a construction-general solution to the problem of the retreat from verb argument structure overgeneralization. *PloS one*, *10*(4), e0123723.

Ambridge, B., Kidd, E., Rowland, C. F., & Theakston, A. L. (2015b). The ubiquity of frequency effects in first language acquisition. *Journal of child language*, *42*(2), 239-273.

Ambridge, B., & Lieven, E. (2015). A constructivist account of child language acquisition. *The handbook of language emergence*, 478-510.

Ambridge, B., Maitreyee, R., Tatsumi, T., Doherty, L., Zicherman, S., Pedro, P. M., Bannard, C., Samanta, S., McCauley, S., Arnon, I., Bekman, D., Efrati, A., Berman, R., Narasimhan, B., Sharma, D. M., Nair, R. B., Fukumura, K., Campbell, S., Pye, C., Fabiola, S., Pixabaj, C., Pelíz, M. M., & Mendoza, M. J. (2020). The crosslinguistic acquisition of sentence structure: Computational modeling and grammaticality judgments from adult and child speakers of English, Japanese, Hindi, Hebrew and K'iche'. *Cognition*, *202*, 104310.

Ambridge, B., Noble, C. H., & Lieven, E. V. (2014). The semantics of the transitive causative construction: Evidence from a forced-choice pointing study with adults and children. *Cognitive Linguistics*, *25*(2), 293-311.

Ambridge, B., Pine, J. M., Rowland, C. F., Chang, F., & Bidgood, A. (2013). The retreat from overgeneralization in child language acquisition: Word learning, morphology, and verb argument structure. *Wiley Interdisciplinary Reviews: Cognitive Science*, *4*(1), 47-62.

Ammon, M. S. H. (1980). *Development in the linguistic expression of causal relations: Comprehension of features of lexical and periphrastic causatives* (Doctoral dissertation, University of California, Berkeley).

Arnold J., Wasow T., Losongco A. & Ginstrom R. (2000). Heaviness vs. Newness: The Effects of Structural Complexity and Discourse Status on Constituent Ordering. *Language,* 17/1, 28-55.

Arterberry, M. E., & Bornstein, M. H. (2001). Three-month-old infants' categorization of animals and vehicles based on static and dynamic attributes. *Journal of experimental child psychology*, *80*(4), 333-346.

Baker, R. K., Pettigrew, T. L., & Poulin-Dubois, D. (2014). Infants' ability to associate motion paths with object kinds. *Infant Behavior and Development*, *37*(1), 119-129.

Barlow, M. & Kemmer, S. (eds). 2000. Usage-based Models of Language. Stanford CA: CSLI.

Barnbrook, G. 1996. Language and Computers. Edinburgh: EUP

Baron, N. S. (1977). Language acquisition and historical change. New York: North-Holland.

Bates, E., & MacWhinney, B. (1989). Functionalism and the competition model. In B. MacWhinney, & E. Bates (Eds.), *The cross-linguistic study of sentence processing*. Cambridge: Cambridge University Press.

Batinić Angster, M. (in press). *Monoclausal configuration with infinitives in Croatian*. Zadar: Sveučilište u Zadru/morepress.

Behrens, H. (2021). Constructivist approaches to first language acquisition. *Journal of Child Language*, *48*(5), 959-983.

Bergs, A. (2020). Complements and Adjuncts. *The Handbook of English Linguistics*, 145-162.

Biber, D. (2010). Corpus-based and corpus-driven analyses of language variation and use. In B., Heine & H. Narrog (Eds.), *The Oxford handbook of linguistic analysis*. Oxford University Press

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (2000). *Longman grammar of spoken and written English*. London: Longman.

Bidgood, A., Pine, J., Rowland, C., Sala, G., Freudenthal, D., & Ambridge, B. E. N. (2021). Verb argument structure overgeneralisations for the English intransitive and transitive constructions: grammaticality judgments and production priming. *Language and Cognition*, *13*(3), 397-437.

Blake, J., Quartaro, G., & Onorati, S. (1993). Evaluating quantitative measures of grammatical complexity in spontaneous speech samples. *Journal of Child Language*, 20(1), 139-152.

Blake, J., Quartaro, G., & Onorati, S. (1993). Evaluating quantitative measures of grammatical complexity in spontaneous speech samples. *Journal of Child Language*, 20(1), 139-152.

Bloom, L., Miller, P., & Hood, L. (1975). Variation and reduction as aspects of competence in language development. In A. Pick (ed.), *Minnesota symposia on child psychology*, vol. IX (3–55). Minneapolis: University of Minnesota Press.

Bloomer, A., & Wray, A. (2006). *Projects in Linguistics: A Practical Guide to Researching Language*. Hodder Arnold.

Boas, H. C. (2013). Cognitive Construction Grammar. In T. Hoffmann & G. Trousdale (eds.), *The Oxford Handbook of Construction Grammar*, 233-254. Oxford: Oxford University Press

Boas, H. C. (2021). Construction grammar and frame semantics. *The Routledge handbook of cognitive linguistics*, 43-77.

Bolinger, D. L. (1968). Entailment and the Meaning of Structures. *Glossa* 2: 119-127.

Booij, G. (2010). *Construction Morphology*. Oxford: Oxford University Press.

Bowerman, M. (1982). Reorganizational processes in lexical and syntactic development. In Wanner & Gleitman (Eds.), *Language acquisition: The state of the art*. Cambridge: Cambridge University Press.

Bowerman, M. (1988). The "no negative evidence" problem. How do children avoid constructing an overgeneral grammar? In J. A. Hawkins (Ed.), *Explaining language universals*. Oxford: Blackwell.

Bowerman, M. (1990). Mapping Thematic Roles onto Syntactic Functions: Are Children Helped by Innate Linking Rules? *Linguistics* 28: 1253-1289.

Braine, M., & Brooks, P. (1995). Verb–argument structure and the problem of avoiding an overgeneral grammar. In M. Tomasello & W. E. Merriman (Eds.), *Beyond names of things: Young children 's acquisition of verbs* (pp. 353 – 376). Hillsdale, NJ: Erlbaum.

Brandt, S., Diessel, H., & Tomasello, M. (2008). The acquisition of German relative clauses: A case study. *Journal of Child Language*, *35*(2), 325-348.

Bresnan, J., & Nikitina, T. (2003). The gradience of the dative alternation. Reality exploration and discovery: Pattern interaction in language and life, 161-184. *On-line*, Stanford University.

Bresnan, J., Cueni, A., Nikitina, T., & Baayen, R. H. (2007). Predicting the dative alternation. In *Cognitive foundations of interpretation* (pp. 69-94). KNAW.

Broccias, C., & Torre, E. (2020). *Attraction and differentiation in the history of the English dative and benefactive alternations.* The Diachrony of Ditransitives, 351, 169 - 195.

Brooks, P. J., & Tomasello, M. (1999). How children constrain their argument structure constructions. *Language*, 720-738.

Brooks, P. J., & Zizak, O. (2002). Does preemption help children learn verb transitivity?. *Journal of Child Language*, *29*(4), 759-781.

Brown, R. (2013). A first language. In *A First Language*. Harvard University Press.

Brown, M., Savova, V., & Gibson, E. (2012). Syntax encodes information structure: Evidence from on-line reading comprehension. *Journal of Memory and Language,* 66(1), 194–209.

Buckle, L., Lieven, E., & Theakston, A. L. (2017). The effects of animacy and syntax on priming: A developmental study. *Frontiers in psychology, 8,* 2246.

Bybee, J. L., & Beckner, C. (2010). Usage-based theory. In B., Heine & H. Narrog (Eds.), *The Oxford handbook of linguistic analysis*. Oxford University Press

Cameron-Faulkner, T., et al., (2007). What part of no do children not understand? A usage-based account of multiword negation. *Journal of Child Language*, 34, 251–282.

Cameron-Faulkner, T., Lieven, E., & Tomasello, M. (2003). A construction based analysis of child directed speech. *Cognitive science,* 27(6), 843-873.

Campbell, A. L., & Tomasello, M. (2001). The acquisition of English dative constructions. *Applied psycholinguistics*, *22*(2), 253-267.

Campbell, A. L., & Tomasello, M. (2001). The acquisition of English dative constructions. *Applied psycholinguistics*, *22*(2), 253-267.

Cann, R., Kaplan, T., & Kempson, R. (2005). Data at the grammar-pragmatics interface: the case of resumptive pronouns in English. *Lingua*, *115*(11), 1551-1577.

Carnerero, C. A. (2007). On "Dimonotransitive" Structures in English, *Conference on Ditransitive Constructions*, 23-25 November 2007, Leipzig (Germany), Max Planck Institute for Evolutionary Anthropology.

Carroll, J. B. (1964). *Language and thought*. Englewood Cliffs, NJ: Prentice-Hall.

Casenhiser, D., & Goldberg, A. E. (2005). Fast mapping between a phrasal form and meaning. *Developmental Science*, 8(6), 500-508.

Chalker, S., & Weiner, E. (2003). Nominal relative clause. *Oxford Reference*. Retrieved 16 Jan. 2020, from https://www.oxfordreference.com/view/10.1093/acref/9780192800879.001.0001/ acref-9780192800879-e-929.

Chomsky, N. (1982). *Some concepts and consequences of the theory of government and binding*. MIT press.

Clark, E. (1987). The principle of contrast: A constraint on language acquisition. In B. MacWhinney (Ed.), *Mechanisms of language acquisition*. 1-33. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Clark, E. V. (2009). *First language acquisition*. Cambridge University Press.

Clark, H. H. (1973). Space, time, semantics, and the child. In *Cognitive development and acquisition of language* (pp. 27-63). Academic Press.

Clifton, C. Jr., & Frazier, L. (2004). Should given information come before new? Yes and no. *Memory & Cognition,* 32, 886–895.

Collins, P. (1995). The indirect object construction in English: an informational approach. *Linguistics, 33(1),* 35-50.

Comrie, B. (1976). The syntax of causative constructions: cross-language similarities and divergences. In *The grammar of causative constructions* (pp. 259-312). Brill.

Corrigan, R. (2012). Using the CHILDES Database in Hoff (Ed.), *Research methods in child language: A practical guide* (Vol. 9). John Wiley & Sons.

Covington, M. A., & McFall, J. D. (2010). Cutting the Gordian knot: The moving-average type–token ratio (MATTR). *Journal of quantitative linguistics*, *17*(2), 94-100.

Croft, W. (2007). Construction grammar. In D. Geeraerts & H. Cuyckens (Eds.), *The Oxford handbook of cognitive linguistics*. Oxford University Press

Crystal, D. (1980). *A first dictionary of linguistics and phonetics.* Boulder, CO: Westview.

Cuetos, F., Mitchell, D. C., & Corley, M. M. (2013). Parsing in different languages. In *Language processing in Spanish* (pp. 163-208). Psychology Press.

Culicover, P. W. (2011). A reconsideration of English relative constructions. *Constructions*, *7*.

Dahl, Ö. (2004). *The growth and maintenance of linguistic complexity.* John Benjamins Publishing Company.

Daller, H., Van Hout, R., & Treffers-Daller, J. (2003). Lexical richness in the spontaneous speech of bilinguals. *Applied linguistics*, *24*(2), 197-222.

Dawson, N., Hsiao, Y., Banerji, N., Tan, A. W. M., & Nation, K. (2021). Features of lexical richness in children's books: Comparisons with child-directed speech. *Language Development Research.*

DeCapua, A. (2008). *Grammar for teachers*. New York: College of New Rochelle.

Devescovi, A., Caselli, M. C., Marchione, D., Pasqualetti, P., Reilly, J., & Bates, E. (2005). A crosslinguistic study of the relationship between grammar and lexical development. *Journal of Child Language,* 32(4), 759-786.

De Vries, M. (2018). Relative clauses in syntax. *Oxford Research Encyclopedia of Linguistics*.

Diessel, H. (2004). *The acquisition of complex sentences* (Vol. 105). Cambridge University Press.

Diessel, H. (2009a). On the role of frequency and similarity in the acquisition of subject and non-subject relative clauses. In T. Givón and M. Shibatani (eds.), Syntactic complexity, 251-276. Amsterdam: John Benjamins.

Diessel, H. (2009b). The emergence of relative clauses in early child language. *MS. University of Jena*.

Diessel, H. (2017). Usage-based linguistics. In *Oxford research encyclopedia of linguistics*.

Diessel, H. & Tomasello, M. (2000). The development of relative clauses in spontaneous child speech. *Cognitive Linguistics* 11: 131-151.

Diessel, H., & Tomasello, M. (2005). A new look at the acquisition of relative clauses. *Language*, 882-906.

Dik, S. C. (1997). *The theory of functional grammar: the structure of the clause*. Walter de Gruyter.

Dixon, R. (1991). *A new approach to English grammar, on semantic princip*les. Oxford: Oxford University Press.

Dixon, R. M. (2000). A typology of causatives: form, syntax and meaning. *Changing valency: Case studies in transitivity*, 30-83.

Dixon, R. M., & Aikhenvald, A. Y. (Eds.). (2000). *Changing valency: Case studies in transitivity* (p. 1). Cambridge: Cambridge University Press.

Durán, P., Malvern, D., Richards, B., & Chipere, N. (2004). Developmental trends in lexical diversity. *Applied Linguistics*, *25*(2), 220-242.

Durrant, P., & Brenchley, M. (2019). Development of vocabulary sophistication across genres in English children's writing. *Reading and writing*, *32*(8), 1927-1953.

Edwards, D., & Goodwin, R. (1986). Action words and pragmatic function in early language. *The development of word meaning: Progress in cognitive development research*, 257-273. New York: Springer-Verlag.

Engber, C. A. (1995). The relationship of lexical proficiency to the quality of ESL compositions. *Journal of Second Language Writing*, 4, 139–155.

Evans, V. (2012). Cognitive linguistics. *Wiley Interdisciplinary Reviews: Cognitive Science*, *3*(2), 129-141.

Fabb, N. (1990). The difference between English restrictive and nonrestrictive relative clauses. *Journal of linguistics*, 26(1), 57-77.

Felser, C., Marinis, T., & Clahsen, H. (2003). Children's processing of ambiguous sentences: A study of relative clause attachment. *Language Acquisition*, *11*(3), 127-163.

Fennell, C. T. (2012).Habituation Procedures in Hoff (Ed.), *Research methods in child language: A practical guide* (Vol. 9). John Wiley & Sons.

Fillmore, C. J. (1975). An alternative to checklist theories of meaning. In *Annual Meeting of the Berkeley Linguistics Society* (Vol. 1, pp. 123-131).

Fillmore, C. J. (1988). The mechanisms of "construction grammar". In *Annual Meeting of the Berkeley Linguistics Society* (Vol. 14, pp. 35-55).

Finney, M. C., Montgomery, J. W., Gillam, R. B., & Evans, J. L. (2014). Role of working memory storage and attention focus switching in children's comprehension of spoken object relative sentences. *Child Development Research*, 450734.

Fischer, S. D. (1972). *The acquisition of verb-particle and dative constructions* (Doctoral dissertation, Massachusetts Institute of Technology).

Fish, G. T. (1968). The indirect object and the redundant construction. *Hispania*, 51(4), 862-866.

Fodor, J. A. (1970). Three reasons for not deriving" kill" from" cause to die". *Linguistic inquiry*, *1*(4), 429-438.

Fodor, J. A. (1983). *The Modularity of Mind.* Cambridge, MA: MIT Press.

Fodor, J. D., & Crain, S. (1984). Simplicity and generality of rules in language acquisition. *Mechanisms of language acquisition*, 35-63.

Ford, M. (1983). A method for obtaining measures of local parsing complexity throughout sentences. *Journal of Verbal Learning and Verbal Behavior*, 22, 203–218.

Friederici, A. D., Fiebach, C. J., Schlesewsky, M., Bornkessel, I. D., & Von Cramon, D. Y. (2006). Processing linguistic complexity and grammaticality in the left frontal cortex. *Cerebral cortex*, 16(12), 1709-1717.

Friedmann, N., & Novogrodsky, R. (2004). The acquisition of relative clause comprehension in Hebrew: A study of SLI and normal development. *Journal of Child Language*, 31, 661-681.

Gaeta, Livio (2017). Markedness. *obo* in Linguistics.

Gast, V., & Diessel, H. (2012). The typology of clause linkage: Status quo, challenges, prospects. *Clause linkage in cross-linguistic perspective: Data-driven approaches to cross-clausal syntax*, 1-36.

Gennari, S. P., & MacDonald, M. C. (2008). Semantic indeterminacy in object relative clauses. *Journal of Memory and Language, 58*, 161–187.

Geeraerts, D. (2006). Introduction a rough guide to cognitive linguistics. In *Cognitive linguistics: Basic readings* (pp. 1-28). De Gruyter Mouton.

Geeraerts, D., & Cuyckens, H. (2007). *The Oxford handbook of cognitive linguistics*. Oxford University Press.

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, 68(1), 1-79.

Gilquin, G. (2004). *Corpus-Based Cognitive Study of the Main English Causative Verbs: A Syntactic, Semantic, Lexical and Stylistic Approach.* Doctoral dissertation, UCL-Université Catholique de Louvain.

Gilquin, G. (2006). The Verb Slot in Causative Constructions. Finding the Best Fit. *Constructions* Special Volume 1, 1–46.

Gilquin, G. (2007). Causing oneself to do something: The psychodynamics of causative constructions. *Linguistics in the twenty first century*, 37-46.

Gilquin, G. (2010). *Corpus, cognition and causative constructions* (Vol. 39). John Benjamins Publishing.

Gilquin, G. (2017). A collostruction-based approach to the integrated contrastive model: The idiomaticity of causative constructions in English, French and French learner English. Paper presented at the *Idiomaticity Workshop*. Oslo: University of Oslo.

Gisborne, N., & Trousdale, G. (2008). *Topics in English Linguistics 57: Constructional Approaches to English Grammar*. Berlin and New York: Mouton de Gruyter.

Givon, T. (1985). Function, Structure, and Language Acquisition, In D. I. Siobin. ed. *The Crosslinguistic Study of Language Acquisition*. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Glynn, D. (2010). Synonyme, lexical fields, and grammatical constructions. A study in usage-based cognitive semantics. In *Cognitive foundations of linguistic usage patterns* (pp. 89-118). De Gruyter Mouton.

Goldberg, A. E. (1989). A unified account of the semantics of the English ditransitive. In *Annual Meeting of the Berkeley Linguistics Society* (Vol. 15, pp. 79-90).

Goldberg, A. E. (1992) The inherent semantics of argument structure: The case of the English ditransitive construction. *Cognitive Linguistics*. vol. 3, no. 1, pp. 37-74.

Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*. University of Chicago Press.

Goldberg, A. E. (2002). Surface generalizations: An alternative to alternations. *Cognitive Linguistics* 13–4, 327–356

Goldberg, A. E. (2006). *Constructions at work: The nature of generalization in language*. Oxford University Press Inc., New York.

Goldberg, A. E. (2013a). Argument structure constructions versus lexical rules or derivational verb templates. *Mind & Language*, *28*(4), 435-465.

Goldberg, A. E. (2013b). The emergence of the semantics of argument structure constructions. In *The emergence of language* (pp. 215-230). Psychology Press.

Goldberg, A. E., Casenhiser, D. M., & Sethuraman, N. (2004). Learning argument structure generalizations. *Cognitive Linguistics*, 15(3), 289-316.

Goldberg, A. E., & Casenhiser, D. (2008). Construction learning and second language acquisition. In *Handbook of cognitive linguistics and second language acquisition* (pp. 207-225). Routledge.

Goldberg, A. E., & Jackendoff, R. (2004). The English resultative as a family of constructions. *Language*, 532-568.

Goldberg, A., & Suttle, L. (2010). Construction grammar. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(4), 468-477.

Goodluck, H., & Tavakolian, S. (1982). Competence and processing in children's grammar of relative clauses. *Cognition*, *11*(1), 1-27.

Goodman, J. C., Dale, P. S., & Li, P. (2008). Does frequency count? Parental input and the acquisition of vocabulary. Journal of child language, 35(3), 515-531.

Gordon, P. C., Hendrick, R., & Johnson, M. (2001). Memory Interference during language processing. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 27, 1411–1423.

Gordon, P. C., Hendrick, R., & Johnson, M. (2004). Effects of noun phrase type on sentence complexity. *Journal of Memory and Language*, 51(1), 97–114

Graham, S. A., & Poulin-Dubois, D. (1999). Infants' reliance on shape to generalize novel labels to animate and inanimate objects. *Journal of child language*, *26*(2), 295-320.

Gries, S. T. (2013). *Statistics for linguistics with R: A practical introduction*. Walter de Gruyter.

Gries, S.T. & Newman, J. (2013). Creating and using corpora in Podesva, R. J. & Sharma, D. (Eds.), *Research methods in linguistics*. Cambridge University Press.

Gries, S. T., & Stefanowitsch, A. (Eds.). (2006). *Corpora in cognitive linguistics: corpus-based approaches to syntax and lexis* (Vol. 172). Walter de Gruyter.

Gropen, J., Pinker, S., Hollander, M., Goldberg, R., & Wilson, R. (1989). The learnability and acquisition of the dative alternation in English. *Language*, 203-257.

Grosu, A. (2012). Towards a more articulated typology of internally headed relative constructions: The semantics connection. *Language and Linguistics Compass*, *6*(7), 447-476.

Guiraud, P. (1960). *Problèmes et méthodes de la statistique linguistique*. Presses universitaires de France.

Guy, G. R., & Bayley, R. (1995). On the choice of relative pronouns in English. *American Speech*, *70*(2), 148-162.

Hadley, P. A., McKenna, M. M., & Rispoli, M. (2018). Sentence diversity in early language development: Recommendations for target selection and progress monitoring. *American Journal of Speech-Language Pathology*, *27*(2), 553-565.

Halliday, M. A. K. (1985). *Spoken and written language*. Geelong. Victoria: Deakin University Press.

Hanks, P. (2000). Contributions of lexicography and corpus linguistics to a theory of language performance. In *Proceedings of the Ninth Euralex International Congress* (pp. 3-13).

Harley H. (2002). Possession and the double-object construction. *In Linguistic Variation Yearbook* 2, ed. P Pica, J Rooryck, pp. 31-70. Amsterdam: Benjamins

Harley, B., & King, M. L. (1989). Verb lexis in the written compositions of young L2 learners. *Studies in Second Language Acquisition*, 11, 415-440.

HarperCollins. (2021). Periphrastic. In *Collins English dictionary*. Retrieved February 15, 2021.

Hart, B., & Risley, T. R. (2003). The early catastrophe: The 30 million word gap by age 3. *American educator*, 27(1), 4-9.

Haspelmath, M. (1997). *From space to time*: *Temporal Adverbials in the World's Languages.* Munich & Newcastle: Lincom Europa.

Haspelmath, M. (2015). Ditransitive constructions. *Annu. Rev. Linguist.*, *1*(1), 19-41.

Heylen, K., Tummers, J., & Geeraerts, D. (2008). Methodological issues in corpus-based Cognitive Linguistics. In *Cognitive Sociolinguistics* (pp. 91-128). De Gruyter Mouton.

Hilpert, M. (2014). Collostructional analysis. *Corpus methods for semantics: Quantitative studies in polysemy and synonymy*, *43*, 391.

Hilpert, M. (2020). Information Structure. *The Handbook of English Linguistics*, 229-248.

Hilpert, M., & Diessel, H. (2017). Entrenchment in construction grammar. In H.-J. Schmid (Ed.), *Entrenchment and the psychology of language learning: How we reorganize and adapt linguistic knowledge* (pp. 57–74). American Psychological Association; De Gruyter Mouton

Hodges, A., Krugler, V., & Law, D. (2004). A corpus study on the item-based nature of early grammar acquisition. *Colorado Research in Linguistics*.

Hoff, E. (2013). *Language development.* Cengage Learning.

Hollmann, W. B. (2003). *Synchrony and Diachrony of English Periphrastic Causatives: A Cognitive Perspective*. University of Manchester.

Hollmann, W. B. (2006). Passivisability of English Periphrastic Causatives. *Corpora in Cognitive Linguistics: Corpus-based approaches to syntax and lexis*, 101-125.

Holmes, V. M., & O'Regan, J. K. (1981). Eye fixation patterns during the reading of relative clause sentences. *Journal of Verbal Learning and Verbal Behavior,* 20, 417–430.

Hovav, M. R., & Levin, B. (2008). The English dative alternation: The case for verb sensitivity. *Journal of linguistics*, *44*(1), 129-167.

Huddleston, R., & Pullum, G. (2002). *The Cambridge grammar of the English language*. Cambridge University Press

Hunston, S., & Francis G. (2000). *Pattern Grammar: A Corpus-Driven Approach to the Lexical Grammar of English*. Amsterdam: Benjamins.

Huttenlocher, J., Vasilyeva, M., Cymerman, E., & Levine, S. (2002). Language input and child syntax. *Cognitive psychology*, *45*(3), 337-374.

Hyltenstam, K. (1988). Lexical characteristics of near-native second-language learners of Swedish. *Journal of Multilingual and Multicultural Development,* 9, 67–84.

Ibbotson, P., & Tomasello, M. (2009). Prototype constructions in early language acquisition. *Language and Cognition*, *1*(1), 59-85.

Jalilevand, N., & Ebrahimipour, M. (2014). Three measures often used in language samples analysis. *Journal of Child Language Acquisition and Development*, 2(1), 1-12.

Jarvis, S. (2013). Capturing the diversity in lexical diversity. *Language Learning*, *63*, 87-106.

Johansson, V. (2008). Lexical diversity and lexical density in speech and writing: A developmental perspective. *Working papers/Lund University, Department of Linguistics and Phonetics*, *53*, 61-79.

Johnson, W. (1944). Studies in language behavior: I. A program of research. *Psychological Monographs*, 56, 1–15.

Jones, G., Cabiddu, F., Barrett, D. J., Castro, A., & Lee, B. (2023). How the characteristics of words in child-directed speech differ from adult-directed speech to influence children's productive vocabularies. *First Language*, 43(3), 253-282.

Jung, Y. J., & Miyagawa, S. (2004). Decomposing ditransitive verbs. In *Proceedings of SICGG* (Vol. 101120).

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, 99, 122–149.

Kastovsky, D. (1973). Causatives. *Foundations of language*, *10*(2), 255-315.

Kecskes, I., & Zhang, F. (2009). Activating, seeking, and creating common ground: A socio-cognitive approach. *Pragmatics & Cognition*, *17*(2), 331-355.

Kidd, E., & Bavin, E. L. (2002). English-speaking children's comprehension of relative clauses: Evidence for general-cognitive and language-specific constraints on development. *Journal of psycholinguistic research*, *31*, 599-617.

Kidd, E., Brandt, S., Lieven, E., & Tomasello, M. (2007). Object relatives made easy: A cross-linguistic comparison of the constraints influencing young children's processing of relative clauses. *Language and cognitive processes*, *22*(6), 860-897.

King, J., & Just, M. A. (1991). Individual differences in syntactic processing: the role of working memory. *Journal of Memory and Language*, 30, 580–602.

Klee, T. (1992). Developmental and diagnostic characteristics of quantitative measures of children's language production. *Topics in Language Disorders*, 12, 28–41.

Kouteva, T., Heine, B., Hong, B., Long, H., Narrog, H., & Rhee, S. (2019). In *World Lexicon of Grammaticalization* (pp. 342–343). H-POSSESSIVE > (4) PERFECT. Cambridge: Cambridge University Press.

Kulikov, L. I., Haspelmath, M., König, E., Oesterreicher, W., & Raible, W. (2001). Causatives. *Language typology and language universals. An international handbook. Vol. 2*, (20/2), 886-898.

Lakoff, G. P. (1966). *On the nature of syntactic irregularity*. Indiana University.

Lakoff, G. (1982). *Categories. An Essay in Cognitive Lingusitics*. University of California

Lakoff, G., & Thompson, H. (1975). Introducing cognitive grammar. In *Annual Meeting of the Berkeley Linguistics Society* (Vol. 1, pp. 295-313).

Langacker, R. W. (1985). Observations and speculations on subjectivity. *Iconicity in syntax*, 109-150. Amsterdam: John Benjamins.

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

Langacker, R. W. (1990). Settings, participants, and grammatical relations. In *Meanings and Prototypes (RLE Linguistics B: Grammar)* (pp. 223-248). Routledge.

Langacker, R. W. (2010). Cognitive Grammar. In B., Heine & H. Narrog (Eds.), *The Oxford handbook of linguistic analysis*. Oxford University Press.

Larson, R. K. (1988). On the double object construction. *Linguistic inquiry, 19*(3), 335-391.

Lauer, S. (2010). Periphrastic Causative Verbs in English: What do They Mean. *Manuscript, Stanford University*.

Leech, G., Rayson, P. & Wilson, A. (2014). *Word frequencies in written and spoken English: Based on the British National Corpus*. Routledge.

Lehmann, C. (1986). On typology of relative clause. *Linguistics* 24 (4). 663–680.

Lemmens, M. (1998). *Lexical perspectives on transitivity and ergativity*. Benjamins.

Levin, B., & Rappaport Hovav, M. (1994). A Preliminary Analysis of Causative Verbs in English. *Lingua*, *92*, 35-77.

Li, H. & Fang, A., C. (2011). Word frequency of the CHILDES corpus: Another perspective of child language features. *ICAME Journal, 35*, 95-116.

Lieven, E. (2010). Input and first language acquisition: Evaluating the role of frequency. *Lingua*, *120*(11), 2546-2556.

Linnarud, M. (1986). *Lexis in composition: A performance analysis of Swedish learners' written English*. Lund, Sweden: CWK Gleerup.

Liu, S., Zhao, X., Li, P. (2008). Early lexical development: A corpus-based study of three languages. In *Proceedings of the 30th Annual Conference of the Cognitive Science* Society (pp. 1343-1348). Austin, TX: Cognitive Science Society.

Logan, K. J., & Conture, E. G. (1995). Length, grammatical complexity, and rate differences in stuttered and fluent conversational utterances of children who stutter. *Journal of Fluency Disorders*, 20(1), 35-61.

Lohmann, A. (2011). Help vs help to: a multifactorial, mixed-effects account of infinitive marker omission1. English Language & Linguistics, 15(3), 499-521.

Louwerse, M. M., McCarthy, P. M., McNamara, D. S., & Graesser, A. C. (2004). Variation in language and cohesion across written and spoken registers. In *Proceedings of the Annual Meeting of the Cognitive Science Society (Vol. 26, No. 26)*.

Lu, X. (2012). The relationship of lexical richness to the quality of ESL learners' oral narratives. *The Modern Language Journal*, *96*(2), 190-208.

Lust, B., Foley, C., & Dye, C. D. (2009). The first language acquisition of complex sentences. *Cambridge Handbook of Child Language*.

MacWhinney, B. (1989). Competition and lexical categorization. *Linguistic categorization*, *61*, 195-241. Amsterdam: John Benjamins.

MacWhinney, B. (2005). Item based constructions and the logical problem. In *Proceedings of the workshop on psychocomputational models of human language acquisition* (pp. 53-68).

MacWhinney, B. (2014). Item-based patterns in early syntactic development. *Constructions, collocations, patterns*, *2562*, 33-69.

MacWhinney, B. (2015). *Tools for Analyzing Talk–Electronic Edition Part 1: The CHAT Transcription Format.* The CHILDES Project: Tools for Analyzing Talk. 3rd Edition. Mahwah, NJ: Lawrence Erlbaum Associates, https://talkbank.org/manuals/CHAT.pdf

MacWhinney, B. (2018). *Tools for Analyzing Talk Part 1: The Chat Transcription Format.* Carnegie Mellon University.

MacWhinney, B. (2023a). *Tools for Analyzing Talk Part 2: The CLAN Program*. Carnegie Mellon University.

MacWhinney, B. (2023b). *Tools for Analyzing Talk Part 3: Morphosyntactic Analysis.* Carnegie Mellon University.

MacWhinney, B., & Pleh, C. (1988). The processing of restrictive relative clauses in Hungarian. *Cognition*, 29, 95–141.

Malchukov, A., Haspelmath, M., & Comrie, B. (2010). Ditransitive constructions: a typological overview. *Studies in ditransitive constructions: A comparative handbook*, *1*, 64.

Malvern, D., Richards, B., Chipere, N., & Durán, P. (2004). *Lexical diversity and language development*. New York: Palgrave Macmillan.

Mandler, J. M., Bauer,P. L. , & McDonough, L. (1991). Separating the sheep from the goats: Differentiating global categories. *Cognitive Psychology*, 23, 263-298.

Marcotte, J. (2005). *Causative Alternation Errors as Event-Driven Construction Paradigm Completions*. Standford University, 1-23.

McKee, C., & McDaniel, D. (2001). Resumptive pronouns in English relative clauses. *Language acquisition*, *9*(2), 113-156.

McKee, C., McDaniel, D. & Snedeker, J. (1998). Relatives Children Say. *J Psycholinguist Res* 27, 573–596.

McMillion, A. (2006). *Labile verbs in English: Their meaning, behavior and structure*. Doctoral dissertation, Engelska institutionen.

Mel'čuk, I. (2021). Relative clause: a typology. *Ten Studies in Dependency Syntax* (pp. 235-274). Berlin, Boston: De Gruyter Mouton.

Miller, J. (2006). Spoken and written English. *The handbook of English linguistics*, 670-691.

Miller, J. F. (1991). Quantifying productive language disorders. In J. F. Miller (Ed.), *Research in child language disorders: A decade of progress* (pp. 211– 220). Austin, TX: Pro-Ed.

Mitchell, D. C., & Cuetos, F. (1991). The origins of parsing strategies. In *Current issues in natural language processing*, 1-12. University of Austin, Texas.

Miyamoto, E., & Nakamura, M. (2003). Subject/Object asymmetries in the processing of relative clauses in Japanese. In G. Garding & M. Tsujimura (Eds.), *Proceedings of WCCFL*, 22, 342-355.

Montag, J. L. (2019). Differences in sentence complexity in the text of children's picture books and child-directed speech. *First language*, *39*(5), 527-546.

Moyle, M. J., Weismer, S. E., Evans, J. L., & Lindstrom, M. J. (2007). Longitudinal relationships between lexical and grammatical development in typical and late-talking children. *Journal of Speech, Language, and Hearing Research*, 50(2): 508–528.

Mukherjee, J. (2004). Corpus data in a usage-based cognitive grammar. In *Advances in Corpus Linguistics* (pp. 83-100). Brill.

Nedjalkov, V. P., & Sil'nickij, G. G. (1973). Typologie der kausativen Konstruktionen. *Folia Linguistica*, *6*(3-4), 273-290.

Ninio, A. (1999). Pathbreaking verbs in syntactic development and the question of prototypical transitivity. *Journal of child language*, *26*(3), 619-653.

Ninio, A. (2003). No verb is an island: Negative evidence on the verb island hypothesis. *Psychology of language and communication*, *7*(1), 3-21.

Núñez, R. (2007). Inferential statistics in the context of empirical cognitive linguistics. *Methods in cognitive linguistics*, 87-118.

Odlin, T. (2012). Nothing will come of nothing. In B. Kortmann & B. Szmrecsanyi (Eds.). *Linguistic complexity: Second language acquisition, indigenization, contact.* Vol. 13 (pp 62-89.). Walter de Gruyter.

Okugiri, M. (2014). English Relative Constructions and Discourse in Spoken and Written Language. *KLA Journal*, (1), 29-40.

O'Grady, W. (2011). Relative clauses: Processing and acquisition. *The acquisition of relative clauses: Processing, typology and function*, 13-38.

O'Loughlin, K. (1995). Lexical density in candidate output on direct and semi-direct versions of an oral proficiency test. *Language testing*, *12*(2), 217-237.

Opfer, J. E., & Gelman, S. A. (2011). Development of the animate-inanimate distinction. *The Wiley-Blackwell handbook of childhood cognitive development*, *2*, 213-238.

Osgood, Charles, and Annette Zehler. 1981. Acquisition of bitransitive sentences: Prelinguistic determinants of language acquisition. Journal of Child Language 8:367-384.ook, Vivian. 1976. A note on indirect objects. Journal of Child Language 3:435-437.

Pallotti, G. (2015). A simple view of linguistic complexity. *Second Language Research*, 31(1), 117-134.

Pavelko, S. L., & Owens Jr, R. E. (2019). Diagnostic accuracy of the Sampling Utterances and Grammatical Analysis Revised (SUGAR) measures for identifying children with language impairment. *Language, Speech, and Hearing Services in Schools*, *50*(2), 211-223.

Pérez-Leroux, A. T. (1995). Resumptives in the acquisition of relative clauses. *Language acquisition*, *4*(1-2), 105-138.

Phillips, J. R. (1973). Syntax and vocabulary of mothers' speech to young children: Age and sex comparisons. *Child development*, 182-185.

Pickering, M. J., & Traxler, M. J. (1998). Plausibility and recovery from garden paths: An eye-tracking study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(4), 940.

Pine, J. M. (1994). The language of primary caregivers. I*nput and interaction in language acquisition*. Eds. C. Gallaway & B.J. Richards.

Pine, J. M., Lieven, E. V., & Rowland, C. F. (1998). Comparing different models of the development of the English verb category. *Linguistics* 36, 807–830.

Pinker, S. (1989) *Learnability and cognition: The acquisition of argument structure*. Cambridge, MA : MIT Press.

Pinker, S. (1994). How could a child use verb syntax to learn verb semantics?. *Lingua, 92*, 377-410.

Pinker, S. (1996). *Language learnability and language development: with new commentary by the author* (1984/1996). Harvard University Press.

Pinker, S. (2013). *Learnability and Cognition, new edition: The Acquisition of Argument Structure*. MIT press.

Proroković, J. & Angster, M. (2022) Kako djeca usvajaju jezik: što nam proizvodnja relativnih rečenica može otkriti o dječjem gramatičkom razvoju?. In Cindrić, M., Ivon, K. & Šimić Šašić, S. (ur.) *Nova promišljanja o djetinjstvu (Zbornik radova)*. Zadar, Sveučilište u Zadru, str. 231-252.

Proroković, J., & Malenica, F. (2023). Language Corpora and First Language Acquisition–A Case Study of the Ditransitive Construction. *Applied Corpus Linguistics, 3*(1), 100041.

Proroković, J. (in preparation). The curious case of the semi-causative *help* construction.

Pye, C., & Levy, Y. (1994). A Cross-Linguistic Approach to the Causative Alternation. *Other Children, Other Languages: Issues in the Theory of Language Acquisition*, 243-63.

Pustejovsky, J. (1995). *The Generative Lexicon*. MIT press.

Race, D. S., & MacDonald, M. C. (2003). The use of "that" in the production and comprehension of object relative clauses. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 25, No. 25).

Rakison, D. H. , & Butterworth, G. E. (1998). Infants' use of object parts in early categorization. *Developmental Psychology*, 34, 49-62 .

Rasinger, S. M. (2013). *Quantitative research in linguistics: An introduction*. A&C Black.

Read, J. (2000). *Assessing vocabulary*. Oxford: Oxford University Press.

Reali, F., & Christiansen, M. H. (2007). Processing of relative clauses is made easier by frequency of occurrence. *Journal of memory and language*, *57*(1), 1-23.

Rice, M. L., Redmond, S. M., & Hoffman, L. (2006). Mean length of utterance in children with specific language impairment and in younger control children shows concurrent validity and stable and parallel growth trajectories. *Faculty Publications, Department of Psychology*. 435.

Rice, M. L., Smolik, F., Perpich, D., Thompson, T., Rytting, N., & Blossom, M. (2010). Mean length of utterance levels in 6-month intervals for children 3 to 9 years with and without language impairments. *Journal of Speech, Language and Hearing Research*, 53(2): 333–349.

Richards, B. (1987). Type/token ratios: What do they really tell us?. *Journal of child language*, *14*(2), 201-209.

Richards, N. (2001). An Idiomatic Argument for Lexical Decomposition. *Linguistic Inquiry* 32, 183-192. MIT Press

Richards, B., Malvern, D. D., Meara, P., Milton, J., & Treffers-Daller, J. (2009). *Vocabulary studies in first and second language acquisition: The interface between theory and application*. Springer.

Robenalt, C., & Goldberg, A. E. (2015). Judgment evidence for statistical preemption: It is relatively better to vanish than to disappear a rabbit, but a lifeguard can equally well backstroke or swim children to shore. *Cognitive Linguistics*, *26*(3), 467-503.

Roeper, T., Lapointe, S., Bing, J., & Tavakolian, S. (1981). A lexical approach to language acquisition. In *Language acquisition and linguistic theory*, ed. Susan Tavakolian, 35-58. Cambridge, Mass.: MIT Press.

Rohdenburg, G. (1996). Cognitive complexity and increased grammatical explicitness in English. *Cognitive Linguistics*, *7*(2), 149-182.

Rondal, J. A., Ghiotto, M., Bredart, S., & Bachelet, J. F. (1987). Age-relation, reliability and grammatical validity of measures of utterance length. *Journal of child language*, *14*(3), 433-446.

Rostad, K., Yott, J., & Poulin-Dubois, D. (2012). Development of categorization in infancy: Advancing forward to the animate/inanimate level. *Infant behavior and development*, *35*(3), 584-595.

Rowe, M. L. (2012). Recording, Transcribing, and Coding Interaction in Hoff (Ed.), *Research methods in child language: A practical guide* (Vol. 9). John Wiley & Sons.

Rusli, Y. A., & Montgomery, J. W. (2017). Children's comprehension of object relative sentences: It's extant language knowledge that matters, not domain-general working memory. *Journal of Speech, Language, and Hearing Research*, *60*(10), 2865-2878.

Sachs, J., Brown, R., i Salerno, R. A. (1976). Adults' speech to children. *Baby talk and infant speech*, 240-245.

Sag, I. A. (1997). English relative clause constructions. *Journal of linguistics*, 33(2), 431-483.

Sánchez-Calderón, S., & Fernández-Fuertes, R. (2018). “Which Came First: The Chicken or the Egg?” Ditransitive and Passive Constructions in the English Production of Simultaneous Bilingual English Children. *Atlantis*, *40*(1), 39-58.

Saffran, J. R., & Thiessen, E. D. (2007). Domain-general learning capacities. In E. Hoff & M. Shatz (Eds.), *Blackwell handbook of language development* (pp. 68–86). Blackwell Publishing.

Saxton, M. (2009). The inevitability of child directed speech. In *Language acquisition* (pp. 62-86). London: Palgrave Macmillan UK.

Scarborough, H., Wyckoff, J., & Davidson, R. A. (1986). Reconsideration of the relation between age and mean utterance length, *Journal of Speech, and Hearing Research*, 29, 394 - 399.

Schäfer, F. (2009). The Causative Alternation. *Language and linguistics compass*, *3*(2), 641-681.

Schmied, J. (1993). Qualitative and quantitative research approaches to English relative constructions. *Corpus based computational linguistics*, 85-96.

Sheldon, A. (1974). *The acquisition of relative clauses in English*. Doctoral dissertation, Indiana University Linguistics Club publication.

Sheldon, A. (1977). On strategies for processing relative clauses: A comparison of children and adults. *Journal of Psycholinguistic Research*, *6*, 305-318.

Shibatani, M. (1976). The grammar of causative constructions: A conspectus. In M. Shibatani (Ed.), *Syntax and semantics: Vol. 6. The grammar of causative constructions* (pp. 1–40). New York: Academic Press.

Shin, G. H. (2022). Automatic analysis of caregiver input and child production: Insight into corpus-based research on child language development in Korean. *Korean Linguistics, 18(2),* 125-158.

Siewierska, A., & Hollmann, W. (2007). Ditransitive clauses in English with special reference to Lancashire dialect. *Structural-functional studies in English grammar*, 83-102.

Slobin, D. I. (1971). Developmental psycholinguistics. *A survey of linguistic science*, 298-411.

Snow, C. E. (1972). Mothers' speech to children learning language. *Child development*, 549-565.

Snyder, W., & Stromswold, K. (1997). The structure and acquisition of English dative constructions. *Linguistic inquiry*, 281-317.

Snyder, W., & Stromswold, K. (1997). The Structure and Acquisition of English Dative Constructions. *Linguistic Inquiry* 28: 281-317.

Snyder, W. (2021). A parametric approach to the acquisition of syntax. *Journal of Child Language, 48*(5), 862-887.

Song, J. J. (1996). *Causatives and Causation. A Universal-Typological Perspective.* London: Longman.

Staub, A., Dillon, B., & Clifton Jr, C. (2017). The matrix verb as a source of comprehension difficulty in object relative sentences. *Cognitive Science*, *41*, 1353-1376.

Stefanowitsch, A. (2001). *Constructing Causation: A Construction Grammar Approach to Analytic Causatives* (Doctoral dissertation).

Stefanowitsch, A. (2011). Cognitive linguistics meets the corpus. *Cognitive linguistics: Convergence and expansion*, 257-290.

Stefanowitsch, A. (2013). Collostructional analysis. In Thomas Hoffmann, and Graeme Trousdale (eds), *The Oxford Handbook of Construction Grammar*. 290–306.

Stefanowitsch, A. (2020). *Corpus linguistics: A guide to the methodology*. Language Science Press.

Stefanowitsch, A., & Gries, S. T. (2003). Collostructions: Investigating the interaction of words and constructions. *International journal of corpus linguistics*, *8*(2), 209-243.

Sugawara, T. (2016). The Syntactic and Semantic Structures of the English Causative to-infinitives. *Interdisciplinary Information Sciences*, 2015-R.

Taeymans, M. (2004). An investigation into the marginal modals dare and need in British present-day English. *Up and down the cline–The nature of grammaticalization*, 97-114.

Talmy, L. (1976). Semantic causative types. In *The grammar of causative constructions* (pp. 41-116). Brill.

Talmy, L. (1978). The relation of grammar to cognition - a synopsis. *American Journal of Computational Linguistics*, 16-26.

Talmy, L. (2000). *Toward a Cognitive Semantics - Volume 1: Concept Structuring Systems* (Vol. 559). Cambridge, MA: mit Press.

Tavakolian, S. L. (1977). *Structural principles in the acquisition of complex sentences*. University of Massachusetts dissertation.

Teubert, W. 2008. Some notes on the concept of cognitive linguistics. In Language, People, Numbers, A. Gerbig & O. Mason (eds), 61–84. Amsterdam: Rodopi

Theakston, A. L. (2004). The role of entrenchment in children's and adults' performance on grammaticality judgment tasks. *Cognitive Development, 19*(1), 15-34.

Thompson, S. A., Longacre, R. E., & Hwang, S. J. J. (1985). Adverbial clauses. *Language typology and syntactic description*, *2*, 171-234.

Tognini-Bonelli, E. (2001). *Corpus linguistics at work* (Vol. 6). John Benjamins Publishing.

Tomasello, M. (1988). The role of joint attentional processes in early language development. *Language sciences*, *10*(1), 69-88.

Tomasello, M. (1992). *First verbs: A case study of early grammatical development*. Cambridge University Press.

Tomasello, M. (1998). Reference: Intending that others jointly attend. *Pragmatics & Cognition*, *6*(1-2), 229-243.

Tomasello, M. (2000a). First steps toward a usage-based theory of language acquisition. *Cognitive Linguistics*, *11*(1-2), 61-82.

Tomasello, M. (2000b). The item-based nature of children's early syntactic development. *Trends in Cognitive Sciences*, *4*(4), 156-163.

Tomasello, M. (2001a). First steps toward a usage-based theory of language acquisition. In *Cognitive Lingusitcs* 11-1/2, 61-82.

Tomasello, M. (2001b). Perceiving intentions and learning words in the second year of life. *Language acquisition and conceptual development*, *3*, 132-158.

Tomasello, M. (2001c). The item-based nature of children's early syntactic development. In M. Tomasello & E. Bates (Eds.), *Essential Readings in Developmental Psychology. Language Development: The Essential Readings* (p. 169–186). Blackwell Publishing.

Tomasello, M. (2005). *Constructing a language: A usage-based theory of language acquisition*. Harvard university press.

Tomasello, M. (2009). *The cultural origins of human cognition*. Harvard university press.

Tomasello, M. (2011). Language development. *The Wiley-Blackwell handbook of childhood cognitive development*, 239-257.

Tomasello, M., & Brooks, P. J. (1999). Early syntactic development: A construction grammar approach. *The development of language*, 161-190.

Tomasello, M., Kuhn, D., & Siegler, R. (2008). Acquiring linguistic constructions. *Child and adolescent development*, 263.

Traugott, E. C., & Trousdale, G. (2014). Contentful constructionalization. *Journal of Historical Linguistics, 4*(2), 256-283.

Traxler, M., Morris, R. K., & Seely, R. E. (2002). Processing subject and object relative clauses: Evidence from eye movements. *Journal of Memory and Language*, 47, 69–90.

Trousdale, G. (2008). Constructions in grammaticalization and lexicalization: Evidence from the history of a composite predicate construction in English. *Topics in English Linguistics 57: Constructional Approaches to English Grammar,* 33-67.

Trudgill, P. (2017). Sociolinguistic Typology: Social Structure and Linguistic Complexity. In A. Aikhenvald & R. Dixon (Eds.), *The Cambridge Handbook of Linguistic Typology* (Cambridge Handbooks in Language and Linguistics, pp. 124-150). Cambridge: Cambridge University Press.

Tsoi, E. Y. L., Yang, W., Chan, A., & Kidd, E. (2019). Mandarin–English speaking bilingual and Mandarin speaking monolingual children's comprehension of relative clauses. *Applied Psycholinguistics, 40*(4), 933-964.

Tummers, J., Heylen, K., & Geeraerts, D. (2005). Usage-based approaches in Cognitive Linguistics: A technical state of the art. Corpus Lingusitics and Linguistic Theory 1(2), 225-261.

Ungerer, T., & Hartmann, S. (2023). *Constructionist approaches: Past, present, future. (Cambridge Elements in Construction Grammar 1).* Cambridge: Cambridge University Press.

Ure, J. (1971). Lexical density and register differentiation. *Applications of linguistics*, *23*(7), 443-452.

Van Dyke, J. A. (2007). Interference effects from grammatically unavailable constituents during sentence processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 33, 407.

Van Dyke, J. A., & Lewis, R. L. (2003). Distinguishing effects of structure and decay on attachment and repair: A cue-based parsing account of recovery from misanalyzed ambiguities. *Journal of Memory and Language*, 49, 285–316.

Whaley, L. J. (1997). *Introduction to typology: The unity and diversity of language*. SAGE publications.

Warren, T., & Gibson, E. (2002). The influence of referential processing on sentence complexity. *Cognition*, 85, 79–112.

Waters, G. S., & Caplan, D. (1996). The capacity theory of sentence comprehension: Critique of Just and Carpenter (1992). *Psychological Review*, 103(4), 761–772.

Wellman, H. M., Harris, P. L., Banerjee, M., & Sinclair, A. (1995). Early understanding of emotion: Evidence from natural language. *Cognition and Emotion*, 9, 117–149.

Wieczorek, R. (2010). Using MLU to study early language development in English. Psychology of Language and Communication, 14(2), 59.

Wolfe-Quintero, K., Inagaki, S., & Kim, H. Y. (1998). *Second language development in writing: Measures of fluency, accuracy, & complexity* (No. 17). University of Hawaii Press.

Wolfe-Quintero, K., Inagaki, S., & Kim, H. Y. (1998). *Second language development in writing : Measures of fluency, accuracy, and complexity* (Report No. 17). Honolulu: University of Hawai'i, Second Language Teaching and Curriculum Center

Wolff, P., & Song, G. (2003). Models of causation and the semantics of causal verbs. *Cognitive psychology*, *47*(3), 276-332.

Wonnacott, E. (2011). Balancing generalization and lexical conservatism: An artificial language study with child learners. *Journal of Memory and Language*, *65*(1), 1-14.

Yoder, P. J. (2006). Predicting lexical density growth rate in young children with autism spectrum disorders. *American Journal of Speech-Language Pathology*, 15, 378–388.

Yoder, P., & Stone, W. L. (2006). A randomized comparison of the effect of two prelinguistic communication interventions on the acquisition of spoken communication in preschoolers with ASD. *Journal of Speech, Language, and Hearing Research*, 49, 698–711

Yoder, P. J., Warren, S. F., & McCathren, R. B. (1998). Determining spoken language prognosis in children with developmental disabilities. *American Journal of Speech Language Pathology*, 7(4), 77–87.

Zenker, F., & Kyle, K. (2021). Investigating minimum text lengths for lexical diversity indices. *Assessing Writing*, *47*, 100505.

Ziegeler, D. (2004). Grammaticalisation through constructions: The story of causative have in English. *Annual Review of Cognitive Linguistics*, 2(1), 159-195.

Zora, S., & Johns-Lewis, C. (1989). Lexical density in interview and conversation. *York Papers in Linguistics*, 14, 89-100.

8.2. CHILDES corpora included in the research:

Following is the list of CHILDES corpora sources used in this research. Though the corpora was accessed via the Sketch Engine platform, any use of it must be accompanied by referencing the work(s) which the authors of the individual subcorpora published, as indicated on the TalkBank pages (see English and American corpora under https://childes.talkbank.org/access/).[103] The access to both English and American corpora according to the author and the name is also available on the following webpage: https://sla.talkbank.org/TBB/childes.

Bates, E., Bretherton, I., & Snyder, L. (1988). *From first words to grammar: Individual differences and dissociable mechanisms*. Cambridge, MA: Cambridge University Press.

Bernstein, N. (1982). *Acoustic study of mothers' speech to language-learning children: An analysis of vowel articulatory characteristics*. Unpublished doctoral dissertation. Bos-ton University.

Bloom, L. (1970). *Language development: Form and function in emerging grammars*. Cambridge, MA: MIT Press

Bohannon, J. N., & Marquis, A. L. (1977). Children's control of adult speech. *Child Development*, 1002–1008.

Brent, M. R. & Siskind, J. M. (2001). The role of exposure to isolated words in early vocabulary development. *Cognition, 81*(2), 31-44.

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

Clark, E. V. (1978a). Awareness of language: Some evidence from what children say and do. In R. J. A. Sinclair & W. Levelt (Eds.), *The child's conception of language.* Berlin: Springer Verlag.

Cruttenden, A. (1978). Assimilation in child language and elsewhere. *Journal of Child Language, 5*, 373–378.

[103] Note that some corpora do not carry the names of the researchers authoring the references listed (e.g. THOMAS, MIAMI, WOLFHEMP etc.). For example, the longitudinal data compiled under the so-called NEW ENGLAND corpus was primarily studied by Catherine Snow and Barbara Pan, with a number of other authors working on the corpus-related publications that are listed as citation information for the use of the corpus on the TalkBank pages.

Davis, Barbara L. & Peter F. MacNeilage (1995). The articulatory basis of babbling. *Journal of Speech and Hearing Research*, 38, 1199-1211.

Demetras, M. (1989a). *Changes in parents' conversational responses: A function of grammatical development.* Paper presented at ASHA, St. Louis, MO.

Demetras, M. (1989b). *Working parents' conversational responses to their two-year-old sons.* University of Arizona.

Demetras, M., Post, K., & Snow, C. (1986). Feedback to first-language learners. *Journal of Child Language, 13*, 275–292.

Dickinson, D. K., & Tabors, P. O. (Eds.) (2001). *Beginning literacy with language: Young children learning at home and school*. Baltimore: Paul Brookes Publishing.

Ervin-Tripp, S. (1978). Some features of early child–adult dialogues. *Language in Society*, *7*(3), 357-373.

Evans, M. A., & Schmidt, F. (1991). Repeated maternal book reading with two children: Language-normal and language-impaired. *First Language*, 11(32), 269-286.

Fletcher, P., & Garman, M. (1988). Normal language development and language impairment: Syntax and beyond. *Clinical Linguistics and Phonetics, 2*, 97–114.

Forrester, M. (2002). Appropriating cultural conceptions of childhood: Participation in conversation. *Childhood, 9,* 255-278.

Garvey, C. (1979). An approach to the study of children's role play. *The Quarterly News-letter of the Laboratory of Comparative Human Cognition*, 12.

Gathercole, Virginia C. (1986). The acquisition of the present perfect: Explaining differences in the speech of Scottish and American children. *Journal of Child Language* 13, 537-560.

Gleason, J. B. (1980). The acquisition of social speech and politeness formulae. In H. Giles, W. P. Robinson, & S. M. P. (Eds.), *Language: Social psychological perspectives*. Oxford: UK: Pergamon.

Gopnik, M. (1989). Reflections on challenges raised and questions asked. In P. R. Zelazo & R. G. Barr (Eds.), *Challenges to developmental paradigms*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Haggerty, L. (1929). What a two-and-one-half-year-old child said in one day. *Journal of Genetic Psychology, 38*, 75–100.

Hall, W. S., Nagy, W. E., & Linn, R. (1984). *Spoken words: Effects of situation and social group on oral word usage and frequency*. Hillsdale, NJ: Erlbaum.

Hayes, D. P. (1986a). The Cornell Corpus. *Technical Report Series 86-1*. Ithaca, NY: Department of Sociology, Cornell University.

Hayes, D. P., & Ahrens, M. G. (1988). Vocabulary simplification for children: A special case of 'motherese'?. *Journal of child language*, *15*(2), 395-410.

Henry, A. (1995). *Belfast English and Standard English: Dialect variation and parameter setting*. New York: Oxford University Press.

Hicks, D. (1990). Kinds of texts: Narrative genre skills among children from two communities. In A. McCabe (Ed.), *Developing narrative structure*. Hillsdale, NJ: Erlbaum.

Higginson, R. P. (1985). *Fixing-assimilation in language acquisition*. Unpublished doctoral dissertation. Washington State University.

Howe, C. (1981). *Acquiring language in a conversational context*. New York: Academic Press.

Inkelas, Sharon & Yvan Rose. 2007. Positional Neutralization: A Case Study from Child Language. *Language* 83, 707-736.

Jones, G., & Rowland, C. F. (2017). Diversity not quantity in caregiver speech: Using computational modeling to isolate the effects of the quantity and the diversity of the input on vocabulary growth. *Cognitive Psychology, 98,* 1-21.

Korman, M. (1984). Adaptive aspects of maternal vocalizations in differing contexts at ten weeks. *First Language,* 5:44-45.

Kuczaj, S. A., & Maratsos, M. P. (1975). What children can say before they will. *Merrill-Palmer Quarterly,* 21, 89–111.

Lieven, E., Salomo, D. & Tomasello, M. (2009). Two-year-old children's production of multiword utterances: A usage-based analysis. *Cognitive Linguistics,20, 3,* 481-508.

Lieven, E., Salomo, D. & Tomasello, M. (2009). Two-year-old children's production of multiword utterances: A usage-based analysis. *Cognitive Linguistics, 20,* 3, 481-508.

Lust, B., & Blume, M. (2016). *Research methods in language acquisition: Principles, procedures, and practices.* Walter de Gruyter GmbH & Co KG.

MacWhinney, B. (1991). *The CHILDES project: Tools for analyzing talk*. Hillsdale, NJ: Erlbaum.

MacWhinney, B., & Bates, E. (1978). Sentential devices for conveying givenness and newness: A cross-cultural developmental study. *Journal of Verbal Learning and Verbal Behavior, 17*, 539–558.

MacWhinney, B., & Snow, C. (1990). The Child Language Data Exchange System: An update. *Journal of Child Language,* 17, 457-472.

McCune, L. (1995). A normative study of representational play at the transition to language. *Developmental Psychology 31*(2), 198-206.

Miranda, E., Camp, L., Hemphill, L., & Wolf, D. (1992). *Developmental changes in children's use of tense in narrative.* Paper presented at the Boston University Conference on Language Development, Boston.

Morisset, C. E., Barnard, K. E., & Booth, C. L. (1995). Toddlers' language development: Sex differences within social risk. *Developmental Psychology, 31*(5), 851-865.

Nelson, K. (Ed.) (1989). *Narratives from the crib.* Cambridge, MA: Harvard University Press.

Ninio, A., Snow, C., Pan, B., & Rollins, P. (1994). Classifying communicative acts in children's interactions. *Journal of Communications Disorders, 27*, 157-188.

Pearson, Barbara Z. (2002). Narrative competence among monolingual and bilingual school children in Miami. In D. K. Oller and R. E. Eilers (Eds.), *Language and literacy in bilingual children* (pp. 135-174). Clevedon, UK: Multilingual Matters.

Peters, A. (1987). The role of imitation in the developing syntax of a blind child. *Text, 7*, 289–311.

R. A. Berman & D. I. Slobin (1994). *Relating events in narrative: A crosslinguistic developmental study.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Rollins, P. R., (2003). Caregiver contingent comments and subsequent vocabulary Comprehension. *Applied Psycholinguistics. 24*, 221-234

Sachs, J. (1983). Talking about the there and then: The emergence of displaced reference in parent–child discourse. In K. E. Nelson (Ed.), *Children's language, Vol. 4*, Hillsdale, NJ: Lawrence Erlbaum Associates.

Soderstrom, M., Blossom, M., Foygel, R., & Morgan, J. L. (2008). Acoustical cues and grammatical units in speech to two preverbal infants. *Journal of Child Language, 35*, 869-902.

Suppes, P. (1974). The semantics of children's language. *American Psychologist, 29*, 103–114.

Tager-Flusberg, H., Calkins, S., Nolin, T., Bamberger, T., Anderson, M., & Chandwick-Dias, A. (1990). A longitudinal study of language acquisition in autistic and Down syndrome children. *Journal of Autism and Developmental Disorders, 20*, 1–21.

Tardif, T., Gelman, S. A., & Xu, F. (1999). Putting the "noun bias" in context: A comparison of English and Mandarin. *Child development*, *70*(3), 620-635.

Theakston, A. L., Lieven, E. V. M., Pine, J. M., & Rowland, C. F. (2001). The role of performance limitations in the acquisition of verb-argument structure: an alternative account. *Journal of Child Language, 28*, 127-152.

Valian, V. (1991). Syntactic subjects in the early speech of American and Italian children. *Cognition, 40*, 21–81.

Van Houten, L. (1986). *Role of maternal input in the acquisition process: The communica-tive strategies of adolescent and older mothers with their language learning children*. Paper presented at the Boston University Conference on Language Development, Boston.

Van Kleeck, A., & Beckley-McCall, A. (2002). A Comparison of Mothers' Individual and Simultaneous Book Sharing With Preschool Siblings. *American Journal of Speech-Language Pathology 11 (2).*

Warren-Leubecker, A. (1982). *Sex differences in speech to children*. Unpublished doctoral dissertation. Georgia Institute of Technology.

Weist, R. M. & Zevenbergen, A. (2008). Autobiographical memory and past time reference. *Language Learning and Development, 4*(4), 291 – 308.

Weismer, S. E. (2017). Typical talkers, late talkers, and children with specific language impairment: A language endowment spectrum?. In *Language disorders from a developmental perspective* (pp. 83-101). Psychology Press.

Wells, C. G. (1981). *Learning through interaction: The study of language development*. Cambridge, UK: Cambridge University Press.

9. ABSTRACT

9.1. Abstract (in English)

This doctoral dissertation investigates various developmental parameters in the context of first language acquisition through a comparative triconstructional analysis (ditransitive, causative, and relative constructions) in children whose first language is English. The research is based on corpus analysis, during which the mentioned constructions were extracted and analysed from the English part of the CHILDES corpora. For a comprehensive interpretation and understanding of first language acquisition, intra-linguistic evidence (i.e., language phenomena related to syntax and lexicon) is observed and compared across three age-defined groups (0-3, 4-6, and 18+). Given the specificities of the three observed constructions, the analyses are tailored to their characteristics (such as frequency, complexity, and variability of certain lexemes and structural patterns appearing in these constructions). Simultaneously, they are conducted to allow not only inter-age group comparisons but also inter-constructional comparisons (such as the degree of similarity in the representation and frequency rankings of certain lexical units, as well as measures like structural and lexical complexity, including the mean length of utterance, lexical density, and type-token ratios). The main contribution of this research is the methodologically unique way of monitoring the parallel development of children's speech at the lexical, morphological, and syntactic levels, which analytically and interpretatively unifies the mentioned constructions through the lens of cognitive theories of language acquisition that are primarily based on the ‘usage-based’ model. Among other things, the research results show what can be described as a linguistic version of Pareto's principle, where all three age groups in the production of the observed constructions largely rely on a few lexical units in certain construction slots, and where the majority of their linguistic expression when it comes to said constructions is occupied by a relatively small number of structural patterns. Significant lexical and structural similarities among the same constructions were confirmed at several levels across the observed age groups, along with a more “conservative” language use and a more pronounced concentration of certain items and syntactic patterns with the declining of age. Furthermore, the ‘item-based’ claims about “simpler” constructions being more lexically restricted production-wise than more “complex” constructions was not confirmed; rather, an examination of certain lexical items used in specific construction slots showed that item-based tendencies are present in all constructions across all three age groups.

In conclusion, the data on frequency and lexical and syntactic complexity obtained through comparative analysis of child and adult speech contribute to existing body of research on the nature of language acquisition and development.

Keywords: *language acquisition, corpus-based research, lexical and structural complexity, ditransitive constructions, periphrastic causative constructions, relative constructions*

9.2. Abstract (in Croatian)

Naslov disertacije:

Razvoj leksika i strukturne složenosti kod usvajanja engleskog kao prvog jezika: dvoprijelazne, kauzativne i relativne konstrukcije

Sažetak:

U ovom doktorskom radu se istražuju različiti jezično-razvojni parametri u kontekstu usvajanja prvog jezika u vidu usporedne trokonstrukcijske analize (dvoprijelazne, kauzativne i relativne konstrukcije) kod djece kojoj je prvi jezik engleski. Istraživanje se temelji na korpusnom istraživanju prilikom koje su se ekstrahirale i analizirale spomenute konstrukcije iz engleskog dijela korpusa CHILDES. U svrhu cjelovitog tumačenja i razumijevanja usvajanja prvog jezika promatraju se i uspoređuju unutarjezični dokazi (odnosno jezične pojave vezane za sintaksu i leksik) između tri dobno određene grupe (0-3, 4-6 i 18+). S obzirom na posebnosti triju promatranih konstrukcija, analize su prilagođene njihovim osobitostima (poput učestalosti, složenosti i varijabilnosti određenih leksema i strukturnih obrazaca koji se pojavljuju u spomenutim konstrukcijama), ali istovremeno provedene na način da, osim međudobnih usporedbi, dopuštaju i međukonstrukcijske usporedbe (poput stupnja različitosti u zastupljenosti i hijerarhiji učestalosti određenih leksičkih jedinica, te mjera poput strukturne i leksičke složenosti, uključujući prosječan broj morfema na sto rečenica, leksičku gustoću, te omjer različnica i pojavnica). Glavni doprinos ovog istraživanja je metodološki jedinstveno praćenje paralelnog razvoja dječjeg govora na leksičkoj, morfološkoj i sintaktičkoj razini, koje analitički i interpretativno objedinjuje navedene konstrukcije kroz prizmu kognitivističkih teorija o usvajanju jezika temeljenima prvenstveno na uporabi (eng. *usage-based model*). Između ostalog, rezultati istraživanja pokazuju ono što se može opisati kao lingvistička inačica Paretovog pravila, gdje se sve tri dobne skupine u proizvodnji promatranih konstrukcija uglavnom oslanjaju na nekolicinu leksičkih jedinica u određenim konstrukcijskim položajima, te da najveći dio njihovog jezičnog izričaja kod proizvodnje istih zauzima relativno mali broj strukturnih obrazaca. Na nekoliko razina je potvrđena značajna leksička i strukturna sličnost kod istih konstrukcija među dobno različitim skupinama, ali i „konzervativnija“ upotreba jezika te izraženija koncentracija određenih elemenata i sintaktičkih uzoraka s padom dobi. Osim toga,

hipoteza da će „jednostavnije“ konstrukcije biti orijentirane na manji broj leksičkih elemenata (eng. *item-based*) od složenijih konstrukcija nije potvrđena; odnosno, pregled određenih leksičkih elemenata korištenih na određenim konstrukcijskim položajima pokazao je da su item-based tendencije prisutne u svim konstrukcijama i dobno različitim skupinama. U konačnici, podatci o učestalosti i leksičko-sintaktičkoj složenosti dobiveni komparativnom analizom govora djece i odraslih doprinose postojećim raspravama o prirodi jezičnog usvajanja i razvoja.

Ključne riječi: *jezično usvajanje, korpusno istraživanje, leksička i jezična složenost, dvoprijelazne konstrukcije, kauzativne konstrukcije, relativne konstrukcije*

10. APPENDIX

10.1. Queries used to target constructions in the corpus

English tagset legend[104]

N.* – all forms of nouns (singular, plural, proper, possessive…)

V.* – all forms of verbs (regardless of tense and number)

PP.* – all forms of pronouns (regular and possessive)

J.* – all forms of adjectives (base form, comparative, superlative)

DT – determiner

PP – personal pronoun

PPZ – possessive pronoun

POS – possessive ending

VV – base form of the verb

VVP – verb, present, not 3rd person

VVD – verb, past tense

VVN – verb, past participle

VVG – verb, gerund/present participle

Example of the tags:

[lemma="have"] – finds all uses of the word have in the corpus

[tag="VVG"] – finds all uses of the gerund in the corpus

DITRANSITIVE CONSTRUCTIONS

V-N-N

[tag="V.*"][tag="N.*"][tag="N.*"] (e.g. *X told James something*)

V-PRO-N

[tag="V.*"][tag="PP"][tag="N.*"] (e.g. *X told him something*)

[104] For a more detailed outline of the English list of part-of-speech tags visit the following Sketch Engine webpage: English Penn Treebank tagset with modifications | Sketch Engine

V-N-PRO

[tag="V.*"][tag="N.*"][tag="PP"] (e.g. *X gave James it*)

V-PRO-PRO

[tag="V.*"][tag="PP"][tag="PP"] (e.g. *X gave him it*)

V-N-MOD-N

[tag="V.*"][tag="N.*"][tag="DT"] (e.g. *X gave James the Y*)

[tag="V.*"][tag="N.*"][tag="J.*"] (e.g *X gave James good Y*)

V-PRO-MOD-N

[tag="V.*"][tag="PP"][tag="DT"] (e.g. *X gave him the Y*)

[tag="V.*"][tag="PP"][tag="J.*"] (e.g. *X gave him good Y*)

V-MOD-N-N

[tag="V.*"][tag="DT"][tag="N.*"][tag="N.*"] (e.g. *X gave the man something*)

[tag="V.*"][tag="J.*"][tag="N.*"][tag="N.*"] (e.g. *X gave white man something*)

[tag="V.*"][tag="DT"][tag="J.*"][tag="N.*"][tag="N.*"] (e.g. *X gave the white man something*)

V-MOD-N-PRO

[tag="V.*"][tag="J.*"][tag="N.*"][tag="PP"] (e.g. *X gave the man it*)

[tag="V.*"][tag="DT"][tag="J.*"][tag="N.*"][tag="PP"] (e.g. *X gave the white man it*)

V-MOD-N-MOD-N

[tag="V.*"][tag="DT"][tag="N.*"][tag="DT"] (e.g. *X gave the man the Y*)

[tag="V.*"][tag="DT"][tag="N.*"][tag="J.*"] (e.g. *X gave the man good Y*)

[tag="V.*"][tag="J.*"][tag="N.*"][tag="DT|J.*"] (e.g. *X gave white man good Y*)

[tag="V.*"][tag="DT"][tag="J.*"][tag="N.*"][tag="DT"] (e.g. *X gave the white man the Y*)

[tag="V.*"][tag="DT"][tag="J.*"][tag="N.*"][tag="J.*"] (e.g. *X gave the white man good Y*)

PERIPHRASTIC (BICLAUSAL) CAUSATIVE CONSTRUCTIONS[105]

HAVE

***have*-N/MOD-N-V**

[lemma="have"] []{0,3} [tag="N.*"][tag="VV|VVP"] (e.g. *Have (the good old) carmechanic repair it.*)

***have*-ProPers-V**

[lemma="have"] []{0,2} [tag="PP.*"] [tag="VV|VVP"] (e.g. *Have (some of) them do it.*)

***have*-ProDem-V**

[lemma="have"] []{0,2} [tag="DT.*"] [tag="VV|VVP"] (e.g. *Have (one of) those do it.*)

GET

***get*-N/MOD-N-*to*-V**

[lemma="get"] []{0,3} [tag="N.*"] [lemma="to"] [tag="VV|VVP"] (e.g. *Get (the good old) carmechanic to repair it.*)

***get*-ProPers-*to*-V**

[lemma="get"] []{0,2} [tag="PP.*"] [lemma="to"] [tag="VV|VVP"] (e.g. *Get (some of) them to do it.*)

***get*-ProDem-*to*-V**

[lemma="get"] []{0,2} [tag="DT.*"] [lemma="to"] [tag="VV|VVP"] (e.g. *Get (one of) those to do it.*)

MAKE

***make*-N/MOD-N-V**

[lemma="make"] []{0,3} [tag="N.*"][tag="VV|VVP"] (e.g. *Make (the good old) carmechanic repair it.*)

[105] The list does not contain the entirety of queries entered into Sketch Engine but only those which were ultimately taken into consideration and included in the data. In order to check for possible deviations from the syntactic norm, the queries in some cases did not always follow the canonical patterns; however, they were typically excluded from the results as they yielded no sensible data, usually confirming doubts about their absence in everyday communicative settings. For example, the periphrastic causative *get* was also tested for its non-infinitival variation ([lemma="get"] []{0,3} [tag="N.*"][tag="VV|VVP"]; e.g. *Get (the good old) carmechanic repair it*) but was ultimately rejected due to absence of resulting phrases. Also, note that passivized causatives such as *have the bloody thing repaired* were not included as they do not constitute proper biclausal causatives (query: [lemma="have"] []{0,3} [tag="N.*"][tag="VVN|VVD"]).

***make*-ProPers-V**

[lemma="make"] []{0,2} [tag="PP.*"] [tag="VV|VVP"] (e.g. *Make (some of) them do it.*)

***make*-ProDem-V**

[lemma="make"] []{0,2} [tag="DT.*"] [tag="VV|VVP"] (e.g. *Make (one of) those do it.*)

LET

***let*-N/MOD-N-V**

[lemma="let"] []{0,3} [tag="N.*"][tag="VV|VVP"] (e.g. *Let (the good old) carmechanic repair it.*)

***let*-ProPers-V**

[lemma="let"] []{0,2} [tag="PP.*"] [tag="VV|VVP"] (e.g. *Let (some of) them do it.*)

***let*-ProDem-V**

[lemma="let"] []{0,2} [tag="DT.*"] [tag="VV|VVP"] (e.g. *Let (one of) those do it.*)

HELP

***help*-N/MOD-N-V**

[lemma="help"] []{0,3} [tag="N.*"][tag="VV|VVP"] (e.g. *Help (the good old) carmechanic repair it.*)

***help*-ProPers-V**

[lemma="help"] []{0,2} [tag="PP.*"] [tag="VV|VVP"] (e.g. *Help (some of) them do it.*)

***help*-ProDem-V**

[lemma="help"] []{0,2} [tag="DT.*"] [tag="VV|VVP"] (e.g. *Help (one of) those do it.*)

***help*-N/MOD-N-to-V**

[lemma="help"] []{0,3} [tag="N.*"] [lemma="to"] [tag="VV|VVP"] (e.g. *Help (the good old) carmechanic to repair it.*)

***help*-ProPers-*to*-V**

[lemma="help"] []{0,2} [tag="PP.*"] [lemma="to"] [tag="VV|VVP"] (e.g. *Help (some of) them to do it.*)

***help*-ProDem-*to*-V**

[lemma="help"] []{0,2} [tag="DT.*"] [lemma="to"] [tag="VV|VVP"] (e.g. *Help (one of) those to do it.*)

OTHER CAUSATIVE PATTERNS CONSIDERED:

query 1: +infinitive

[lemma="X"] []{0,3} [tag="N.*"| tag="PP.*"| tag="DT.*"] [lemma="to"] [tag="VV|VVP"] (e.g. *Force someone to do something.*)

query 2: +into gerund

[lemma="X"] []{0,3} [tag="N.*"| tag="PP.*"| tag="DT.*"] [lemma="into"] [tag="VVG"] (e.g. *Pressure someone into doing something.*)

CAUSE-NP-*to*-V

X= cause|force|compel|coerce|lead|oblige|drag|drive|press|pressure|blackmail|strong-arm

TELL-NP-*to*-V

X= order|command|appoint|direct|require|request|demand|dictate|ordain|convince|motivate|urge|ask|beg|bid|tell|instruct|authorize|empower|entitle|endorse|sanction|support|insist|talk

ALLOW-NP-*to*-V

X= allow|permit|enable|grant|license|approve

TALK-NP-*into*-V

X= order|command|appoint|direct|require|request|demand|dictate|ordain|convince|motivate|urge|ask|beg|bid|tell|instruct|authorize|empower|entitle|endorse|sanction|support|insist|talk.

RELATIVE CONSTRUCTIONS[106]

NP-*that*-VP

[tag="N.*"] [lemma="that"] [tag="V.*"] (e.g. *The man that is…*)

[tag="PP.*"] [lemma="that"] [tag="V.*"] (e.g. *He that is…*)

[tag="DT.*"] [lemma="that"] [tag="V.*"] (e.g. *Those that are…*)

[106] As with periphrastic causatives, the list does not contain the entirety of queries entered into Sketch Engine but only those which were ultimately retained in the final dataset. For example, one such query was the subject relative with the relativizer *whose*, which is ungrammatical and was ultimately rejected. Consider [tag="N.*"] [lemma="whose"] [tag="V.*"], which would translate to an example like *Man whose knew…*, or [tag="N.*"] [lemma="where"] [tag="V.*"] which translates to an example like *The place where happens…,* both of which may have been accidentally produced in spontaneous speech. Across all three constructions, such idiosyncrasies were so rare (almost non-existent) that there was no need to form a category for them in the results section, and if they were observed, it is often difficult to assess whether the intended utterance by the speaker truly represented the targeted construction.

NP-*that*-NP-VP

[tag="N.*"] [lemma="that"] [tag="J.* "| tag="DT.*"| tag="PPZ"| tag="POS"]{0,2}[tag="PP.*"|tag="DT.*"|tag="N.*"] [tag="V.*"] (e.g. *The place that X used to…*)

[tag="PP.*"] [lemma="that"] [tag="J.* "| tag="DT.*"| tag="PPZ"| tag="POS"]{0,2}[tag="PP.*"|tag="DT.*"|tag="N.*"] [tag="V.*"] (e.g. *It that X used to…*)

[tag="DT.*"] [lemma="that"] [tag="J.* "| tag="DT.*"| tag="PPZ"| tag="POS"]{0,2}[tag="PP.*"|tag="DT.*"|tag="N.*"] [tag="V.*"] (e.g. *There that X used to…*)

NP-*which*-VP

[tag="N.*"] [lemma="which"] [tag="V.*"] (e.g. *The place which used to…*)

[tag="PP.*"] [lemma="which"] [tag="V.*"] (e.g. *It which used to…*)

[tag="DT.*"] [lemma="which"] [tag="V.*"] (e.g. *There which used to…*)

NP-*which*-NP-VP

[tag="N.*"] [lemma="which"] [tag="J.* "| tag="DT.*"| tag="PPZ"| tag="POS"]{0,2}[tag="PP:?"|tag="DT.*"|tag="N.*"] [tag="V.*"] (e.g. *The place which X used to…*)

[tag="PP.*"] [lemma="which"] [tag="J.* "| tag="DT.*"| tag="PPZ"| tag="POS"]{0,2}[tag="PP:?"|tag="DT.*"|tag="N.*"] [tag="V.*"] (e.g. *It which X used to…*)

[tag="DT.*"] [lemma="which"] [tag="J.* "| tag="DT.*"| tag="PPZ"| tag="POS"]{0,2}[tag="PP:?"|tag="DT.*"|tag="N.*"] [tag="V.*"] (e.g. *There which X used to…*)

NP-*who*-VP

[tag="N.*"] [lemma="who|whom"] [tag="V.*"] (e.g. *Man who was…*)

[tag="PP.*"] [lemma="who|whom"] [tag="V.*"] (e.g. *He who was…*)

[tag="DT.*"] [lemma="who|whom"] [tag="V.*"] (e.g. *Those who were…*)

NP-*who*-NP-VP

[tag="N.*"] [lemma="who|whom"] [tag="J.* "| tag="DT.*"| tag="PPZ"| tag="POS"]{0,2}[tag="PP:?"|tag="DT.*"|tag="N.*"] [tag="V.*"] (e.g. *Man whom X knew…*)

[tag="PP.*"] [lemma="who|whom"] [tag="J.* "| tag="DT.*"| tag="PPZ"| tag="POS"]{0,2}[tag="PP:?"|tag="DT.*"|tag="N.*"] [tag="V.*"] (e.g. *Him whom X knew…*)

[tag="DT.*"] [lemma="who|whom"] [tag="J.* "| tag="DT.*"| tag="PPZ"| tag="POS"]{0,2}[tag="PP:?"|tag="DT.*"|tag="N.*"] [tag="V.*"] (e.g. *Those whom X knew*…)

NP-*whose*-NP-VP

[tag="N.*"] [lemma="whose"] [tag="J.* "| tag="DT.*"| tag="PPZ"| tag="POS"]{0,2}[tag="PP:?"|tag="DT.*"|tag="N.*"] [tag="V.*"] (e.g. *Man whose X knew*…)
[tag="PP.*"] [lemma="whose"] [tag="J.* "| tag="DT.*"| tag="PPZ"| tag="POS"]{0,2}[tag="PP:?"|tag="DT.*"|tag="N.*"] [tag="V.*"] (e.g. *He whose X knew*…)
[tag="DT.*"] [lemma="whose"] [tag="J.* "| tag="DT.*"| tag="PPZ"| tag="POS"]{0,2}[tag="PP:?"|tag="DT.*"|tag="N.*"] [tag="V.*"] (e.g. *Those whose X knew*…)

NP-*where*-NP-VP

[tag="N.*"] [lemma="where"] [tag="J.* "| tag="DT.*"| tag="PPZ"| tag="POS"]{0,2}[tag="PP:?"|tag="DT.*"|tag="N.*"] [tag="V.*"] (e.g. *The place where X happens*…)
[tag="PP.*"] [lemma="where"] [tag="J.* "| tag="DT.*"| tag="PPZ"| tag="POS"]{0,2}[tag="PP:?"|tag="DT.*"|tag="N.*"] [tag="V.*"] (e.g. *It where X happens*…)
[tag="DT.*"] [lemma="where"] [tag="J.* "| tag="DT.*"| tag="PPZ"| tag="POS"]{0,2}[tag="PP:?"|tag="DT.*"|tag="N.*"] [tag="V.*"] (e.g. *There where X happens*…)

NP-*when*-NP-VP

[tag="N.*"] [lemma="when"] [tag="J.* "| tag="DT.*"| tag="PPZ"| tag="POS"]{0,2}[tag="PP:?"|tag="DT.*"|tag="N.*"] [tag="V.*"] (e.g. *The time when X happens*…)
[tag="PP.*"] [lemma="when"] [tag="J.* "| tag="DT.*"| tag="PPZ"| tag="POS"]{0,2}[tag="PP:?"|tag="DT.*"|tag="N.*"] [tag="V.*"] (e.g. *It when X happens*…)
[tag="DT.*"] [lemma="when"] [tag="J.* "| tag="DT.*"| tag="PPZ"| tag="POS"]{0,2}[tag="PP:?"|tag="DT.*"|tag="N.*"] [tag="V.*"] (e.g. *Then when X happens*…)

10.2. Additional data on constructions

For a detailed explanation of subsequent measures calculated with the 'eval' function in CLAN (Table 40, Table 41, Table 42, Table 43) see MacWhinney (2023a, pp.132-133).

Table 40. MLUs and other data on constructions; 'eval' function in CLAN

Construction	Age group	Total_Utts	MLU_Utts	MLU_Words	MLU_Morphemes	FREQ_types	FREQ_tokens	FREQ_TTR	Verbs_Utt
ditransitive *give*-NP-NP	0-3	1014	978	6.56	7.16	682	6975	0.10	1.39
	4-6	798	768	8.32	9.21	799	6960	0.12	1.63
	18+	11205	11052	8.47	9.30	3283	99654	0.03	1.66
causative *make*-NP-V	0-3	386	375	8.06	8.59	457	3129	0.15	2.35
	4-6	298	280	10.43	11.25	536	3163	0.17	2.62
	18+	3204	3146	9.15	9.82	1909	29756	0.06	2.43
relative N-*that*-NP-VP	0-3	206	203	10.23	11.29	431	2132	0.20	1.75
	4-6	497	475	13.06	14.09	855	6513	0.13	2.22
	18+	5251	5144	13.59	14.78	3626	71902	0.05	2.23
random sample	0-3	20739	12222	3.89	4.16	3176	69791	0.05	0.50
	4-6	16377	10748	5.24	5.61	3564	77242	0.05	0.72
	18+	18290	12088	4.45	4.75	3853	73818	0.05	0.61

Table 41. Part-of-Speech proportions in constructions; 'eval' function in CLAN

Construction	Age group	%_Nouns	%_Plurals	%_Verbs	%_prep	%_adj	%_adv	%_conj	%_det	%_pro
ditransitive *give*-NP-NP	0-3	18%	9%	21%	2%	2%	2%	1%	8%	29%
	4-6	18%	11%	19%	2%	2%	3%	1%	10%	26%
	18+	18%	8%	19%	3%	3%	4%	1%	10%	23%
causative *make*-NP-V	0-3	11%	13%	31%	3%	2%	7%	1%	5%	23%
	4-6	14%	18%	26%	4%	2%	7%	2%	5%	22%
	18+	12%	15%	28%	4%	3%	6%	2%	5%	22%
relative N-*that*-NP-VP	0-3	17%	27%	19%	4%	3%	5%	1%	10%	24%
	4-6	18%	19%	18%	5%	3%	5%	1%	9%	23%
	18+	18%	19%	18%	7%	3%	5%	1%	10%	20%
random sample	0-3	19%	10%	16%	4%	4%	6%	1%	6%	16%
	4-6	16%	13%	16%	5%	4%	6%	1%	7%	17%
	18+	17%	12%	16%	5%	4%	7%	1%	7%	16%

Table 42. Tenses and forms in constructions; 'eval' function in CLAN

Construction	Age group	%_Aux	%_Mod	%_3S	%_13S	%_PAST	%_PASTP	%_PRESP
ditransitive *give*-NP-NP	0-3	2%	3%	5%	0%	13%	1%	11%
	4-6	1%	3%	11%	1%	17%	1%	10%
	18+	2%	4%	7%	1%	14%	3%	10%
causative *make*-NP-V	0-3	0%	3%	6%	1%	6%	0%	7%
	4-6	1%	2%	9%	1%	13%	1%	7%
	18+	1%	3%	8%	1%	10%	1%	7%
relative N-*that*-NP-VP	0-3	1%	1%	14%	2%	24%	5%	12%
	4-6	1%	1%	12%	5%	25%	3%	9%
	18+	1%	1%	14%	5%	32%	5%	10%
random sample	0-3	1%	2%	14%	1%	12%	4%	13%
	4-6	1%	2%	13%	3%	18%	4%	11%
	18+	1%	2%	13%	2%	13%	4%	10%

Table 43. Open class and closed class words ratios; 'eval' function in CLAN

Construction	Age group	noun_verb	open_closed	#open-class	#closed-class
ditransitive *give*-NP-NP	0-3	0.8	0.8	3008	3860
	4-6	0.9	0.8	2963	3868
	18+	0.9	0.8	43926	53936
causative *make*-NP-V	0-3	0.4	1.1	1629	1454
	4-6	0.5	1.0	1543	1547
	18+	0.4	1.0	14789	14382
relative N-*that*-NP-VP	0-3	0.9	0.8	935	1167
	4-6	1.0	0.8	2816	3571
	18+	1.0	0.8	31498	39437
random sample	0-3	1.1	1.0	32199	33465
	4-6	1.0	0.8	33658	39693
	18+	1.1	0.9	33516	36206

Table 44. Token-based animacy of the nouns in the observed construction slots

Construction	Construction slot	Age group	INAN	ANc	ANp	ANt	ANb
ditransitive *give*-NP-NP	1st (indirect) object	0-3	2%	94%	3%	1%	0%
		4-6	2%	95%	2%	1%	0%
		18+	6%	86%	6%	2%	0%
ditransitive *give*-NP-NP	2nd (direct) object	0-3	94%	3%	1%	1%	1%
		4-6	95%	1%	1%	2%	1%
		18+	96%	1%	0%	1%	2%
causative *make*-NP-V	causee	0-3	28%	39%	19%	10%	4%
		4-6	45%	32%	13%	5%	5%
		18+	37%	37%	9%	9%	9%
relative N-*that*-NP-VP	relativized noun	0-3	89%	9%	0%	1%	0%
		4-6	87%	8%	0%	4%	1%
		18+	90%	6%	1%	2%	1%

For a detailed explanation of indices on lexical variation outlined in the subsequent tables (Table 45, Table 46, Table 47) see Lu (2012).

Table 45. Measures of lexical sophistication

Construction	Age group	LD	LS1	LS2	VS1	VS2	CVS1
ditransitive *give*-NP-NP	0-3	0.44	0.25	0.40	0.00	0.00	0.04
	4-6	0.45	0.23	0.40	0.02	0.12	0.24
	18+	0.45	0.26	0.41	0.03	0.29	0.38
causative *make*-NP-V	0-3	0.51	0.14	0.31	0.03	0.62	0.56
	4-6	0.50	0.14	0.30	0.05	1.30	0.80
	18+	0.52	0.19	0.38	0.05	1.46	0.86
relative N-*that*-NP-VP	0-3	0.42	0.24	0.36	0.05	0.94	0.69
	4-6	0.40	0.20	0.37	0.07	2.01	1.00
	18+	0.41	0.21	0.37	0.05	1.18	0.77

Table 46. Measures of lexical variation

Construction	Age group	NDW	NDW-Z-50	NDW-ER50	NDW-ES50	TTR	MSTTR-50	CTTR	RTTR	LOGTTR	UBER
ditransitive *give*-NP-NP	0-3	257	28	29.70	28.70	0.19	0.56	4.96	7.02	0.77	13.63
	4-6	360	27	31.10	30.60	0.21	0.61	6.20	8.77	0.79	15.53
	18+	379	31	33.80	30.70	0.21	0.64	6.29	8.89	0.79	15.60
causative *make*-NP-V	0-3	321	30	34.80	30.40	0.18	0.63	5.39	7.62	0.77	14.21
	4-6	417	33	32.90	33.10	0.19	0.66	6.30	8.91	0.78	15.49
	18+	427	31	35.00	34.80	0.22	0.68	6.93	9.80	0.80	16.58
relative N-*that*-NP-VP	0-3	412	30	35.40	32.90	0.19	0.64	6.26	8.85	0.78	15.43
	4-6	470	36	36.00	32.60	0.18	0.66	6.46	9.14	0.78	15.61
	18+	604	34	36.10	33.80	0.20	0.70	7.77	10.99	0.80	17.32

Table 47. Measures of lexical variation

Construction	Age group	VV1	SVV1	CVV1	LV	VV2	NV	ADJV	ADVV	MODV
ditransitive *give*-NP-NP	0-3	0.09	2.12	1.03	0.34	0.04	0.63	0.02	0.03	0.06
	4-6	0.15	6.62	1.82	0.38	0.06	0.61	0.04	0.04	0.07
	18+	0.16	8.24	2.03	0.39	0.07	0.62	0.06	0.03	0.09
causative *make*-NP-V	0-3	0.14	10.17	2.25	0.29	0.08	0.70	0.03	0.04	0.06
	4-6	0.19	21.48	3.28	0.31	0.10	0.63	0.03	0.04	0.07
	18+	0.20	20.96	3.24	0.36	0.11	0.74	0.04	0.06	0.10
relative N-*that*-NP-VP	0-3	0.28	26.79	3.66	0.39	0.11	0.51	0.03	0.04	0.07
	4-6	0.27	28.74	3.79	0.37	0.10	0.52	0.04	0.04	0.08
	18+	0.26	30.94	3.93	0.42	0.10	0.61	0.04	0.05	0.09