ACTIONS AND EVENTS

TEMPORAL AND MULTILAYERED STRUCTURES
IN LANGUAGE AND COGNITION

by

Kristian Nicolaisen

A dissertation submitted in partial fulfillment for the requirements of the degree of Doctor of Philosophy

Heidelberg University
2023

Abstract

Perceiving events as they unfold in the world around us involves recognizing features of human actions, relations between event participants, and relations between different states of the event participants as they change over time. This is achieved by associating perceptual experience with generalized memory structures, a mental capacity which is attributed to temporarily maintaining event representations in working memory. Often, these non-linguistic memory structures are captured in terms of their linguistic expression as event structures, sets of thematic roles, schemata, or frames. However, when these concepts form the sole basis for empirical investigation, fine-grained features of event representation are lost, giving rise to challenges for cognitive sciences in providing consistently applicable models. The aim of this dissertation is to provide a theoretical basis for identifying overlooked aspects in event representation and to experimentally test central hypotheses generated from these ideas. Drawing on assumptions laid out in ontology, psychology, and linguistics (especially focusing on theory of human action, event representation, lexical semantics, and Construction Grammar), it is first explored how cognitive representation and verbal expression of actions and events are multilayered in nature: Our understanding of actions and events that coincide in time and space form "thick" representations consisting of layers that we switch our attention between or superimpose. This basic aspect of event representation will be investigated through the concept of *cascades*. Second, since events are defined by the changes their participants undergo over time, their internal structure is temporally complex. This dynamic component of events will be explored through the concepts of *Argument-time Structures* and *Intersecting Object Histories*. In terms of cognitive representation, it is predicted that an event comprises multiple temporal phases containing different states of its event participants, while language provides means of highlighting certain phases. For instance, speakers of English can express different aspectual properties of events, such as whether an event is ongoing

(progressive aspect) or completed (perfect aspect). This idea is tested in the experimental part of the dissertation. By using a modified visual-world-paradigm, grammatical aspect markers on double-object verbs were manipulated to investigate whether and how quickly aspect contributes to the mapping of linguistic form onto temporal properties of transfer-of-possession event representations. Results from Experiment 1 show that listeners use aspect cues to predict and integrate states of referents associated with different temporal event stages. In Experiment 2, the conventional double-object verb from Experiment 1 were replaced by novel verbs in double-object sentence forms to determine whether syntactic structures can be mapped onto event structure without access to lexicalized verbs. This was done to dissociate which components in the sentence (syntactic vs. lexical) can be relied on in event comprehension. Results show that sentence meanings can be derived from syntactic structures in the absence of a lexical verb.

Together, the theoretical analyses and experimental results contribute to an understanding of actions and events that goes beyond simple semantic categorization such as thematic structures or event structures, but which takes a fine-grained approach to the underlying complex cognitive structures. This offers the possibility for future research to investigate in more detail the structures that are mentally instantiated during language processing, both concerning conceptual representation and linguistic encoding.

Acknowledgments

First and foremost, I would like to express my sincerest gratitude to my supervisor Prof. Dr. Christiane von Stutterheim. Ever since I attended her introductory courses, throughout her graduate courses in psycholinguistics and theoretical linguistics, and to the very end of my doctoral project, her insights and unwavering willingness to discuss ideas and provide encouraging support have been an invaluable experience. Without her guidance, clear mind, and enthusiastic person, this project would not have been possible.

I am also deeply thankful to Dr. Johannes Gerwien for sparking my initial interest for event cognition; his commitment to the project and his crucial contribution to the experimental part of this work as well as to my methodological training has been of the greatest value. I owe him a great debt for his creativity and endless discussions on method and ideas. My debt also goes to PD Dr. Sandra Pappert and Dr. Agnieszka Konopka for sharing their thoughts on the work, providing guidance, and for giving me the opportunity to conduct experiments, and to Prof. Dr. Michael Herweg and Prof. Dr. Sebastian Löbner for very valuable input. Many other people at the Heidelberg University Language at Cognition Lab and at institutions abroad have contributed to the development of the project, especially my colleagues, Ines Marberg and Danny Dirker, whom I thank for sharing their insights, questions, and their passion for the field.

I am thankful to the German Academic Scholarship Foundation for allowing me to pursue this project by granting me with a doctoral scholarship and for extending it to compensate for pandemic-related challenges in collecting data.

Finally, I am indebted to my family, Anne, Svend, Frits, Line, and Johannes, and to my friends for providing unconditional encouragement and loving support along the way. I am also immensely grateful to Nina. Even though there has been a great distance between us, she has always been there.

Abbreviations

ATS	Argument-time Structure Theory
AOI	Area of Interest
CM	Caused motion
CP	Caused possession
DO	Double-object
fMRI	Functional magnetic resonance imaging
IA	Interest Area
IP	Interest Period
IOH	Intersecting Object Histories
LOC	Location
OBJ	Object
P	Predicate
PERF	Perfect aspect
PO	Prepositional-object
PROG	Progressive aspect
SUBJ	Subject
t1, 2, 3	time 1, 2, 3

List of tables and figures

Table of Contents

1 Introduction

Through human action, events take place in the world. We hold the door open for someone so that he or she can enter, we play loud music and wake up our neighbor, we give money to the cashier and pay for goods. Behind each of these specific actions and the effects they cause, we can recognize fundamental principles that we experience again and again every day: We act, and in doing so, we cause something to happen.

The epistemological interest of the present work originally arose from reflections on human actions that can be expressed by sentences such as *give me the book* and *tell me a story*. Both sentences count as instances of the double-object construction, a not exactly underexplored linguistic phenomenon. Its meaning – ranging from a literal to a metaphorical sense of transfer of possession – has been scrutinized to such a degree that, according to Goldberg (2019), it has garnered a central role in linguistics, much as the fruit fly has been treated in biology. Yet research on this construction seems to have brought more division than unification. Even among researchers who can agree on central aspects of how its form relates to its meaning, there are fundamental viewpoints that cannot be reconciled. Research has thus left behind contradictory ideas about a basic sentence form that is part of the core grammar for speakers of many languages. If we look beyond the English language, the picture becomes even more diffuse. Consider the following usage of the construction in German: *Die Mutter verbietet ihrem Sohn den Computer* (eng. 'the mother forbids her son the computer'). Does the meaning of this sentence fall within the above-mentioned definition (i.e., transfer of possession), or does it count as evidence that the double-object construction not necessarily expresses any type of transfer of possession? We will return to this example in chapter 4.

In many studies on basic sentence forms, it seems desirable to explain their content in terms of their cognitive representation. However, the more precise nature of the cognitive representations underlying linguistic meaning is rarely investigated. Instead, a few

axiomatic approaches are typically adopted to define such representations, perhaps most notably the thematic structure approach (cf. Fillmore, 1968). This may be a reason for fruitless attempts at a coherent theory which captures the semantic variations of the double-object construction. In this dissertation, it will be attempted to tackle these challenges, and do so by examining the approach to describing cognitive representation more closely. This will be done by focusing on the origin of the formation of cognitive structures – in the mind of the speaker and listener.

Our starting point for approaching this topic can be captured with the following observation: In order to navigate the world, we have to break down a stream of sensory input into comprehensible units that can be stored in our working and long-term memory. We are able to do this by drawing on previous experiences with the world that have consolidated as generalized chunks of knowledge (Zacks et al., 2007). Knowing how a birthday party unfolds, that a traffic light will eventually turn green, or that winter will turn into spring – these are just a few examples. Each of these is associated with the notion *events* in cognitive science research on perception, memory, and language, and in philosophy. The notion is used to capture a dynamic aspect of the world that involves something that we can perceive, remember, and talk about, including any basic understanding of the double-object construction (e.g., transfer-of-possession meaning). This notion will be the primary focus of the present dissertation. Actions are defined as a subtype of events involving human activity, potentially filling gaps in a schematic event structure that represents more general cause-effect relationships. In the case of the different instances of the double-object construction, this means that transfer of possession can take place via the act of giving (which can denote a physical transfer of an object between two people) and the act of telling (which can denote a transfer of information between two people).

Defining what events are varies considerably between scientific disciplines. Even within lines of research concerned with their cognitive representation, there is little clarity about what exactly they consist of. The perhaps most widely adopted approach to the study of the cognitive or linguistic representation of events goes back to the linguistic theory of Fillmore (1968) and Dowty (1989). In this approach, events are captured as semantic structures. These describe our mental ability to classify basic relational

structures, such as *who did what to whom*. If we want to report about the guest who dropped the cake at the birthday party last night, we can conceptualize the guest as the acting role *agent* and the cake as an affected role *patient*, so that we can represent the situation in a format that can be mapped onto words and sentence forms and be articulated (Levelt, 1989; see Konopka & Brown-Schmidt, 2014). However, when one looks beyond the studies that assume this axiomatic semantic approach, the discussion of the basic structure of events becomes much more complicated. This is in part due to the fact that semantic structures only explain an aspect of event representation relevant for linguistic encoding. As a result, the approach lacks core aspects of event structures, one of which is related to their dynamic component, or in terms of cognitive processing: related to our ability to perceive change (or stasis) over time (see Langacker, 1987: p. 258).

In an account that explicitly positioned itself against the conventional semantic approach, Wolfgang Klein captured the dynamic component in the meaning of verbs. Klein took research on event semantics a step further by questioning the generalizability of semantic roles across linguistic phenomena and instead adopting a perspective that considered the internal, logico-temporal structures of events. As his research shows, utterances in many languages (with exceptions such as Chinese) are associated with particular constellations of temporally specific features of event entities (e.g., 'before'- and 'after'-states of acting and affected entities) (Klein, 1999; 2002; 2010). With respect to speech production processes, it follows that bundling of temporal features of event participants takes place during event conceptualization (see Gerwien, 2015).

In the research of Christiane von Stutterheim, Johannes Gerwien, and colleagues, further aspects of the complexity of event representations have been addressed and tested experimentally. The core of this research goes back to the idea in ontology that multiple events can coincide in time and space, forming multilayered or "thick" events (Bennett, 2002). Through this approach, the philosophical debate on the multilayeredness of events has been moved into cognitive science and experimentally tested. Studies by Stutterheim, Gerwien, and colleagues show how directing attention to a particular layer of an event determines how events are divided into units. For example, if we imagine someone running across a train station and boarding a train, we can either perceive a layer with the

path that a person leaves behind, which involves two separate events (‘crossing the station’ and ‘boarding the train’), or we can perceive a layer with a single event that represents the manner of motion exhibited by the person (running). Central to the approach are the findings that languages such as German and French differ in which of these layers is typically expressed by the verb, leading to cross-linguistic variations in event conceptualization (Gerwien & Stutterheim, 2022; Gerwien & von Stutterheim, 2022; Lambert et al., 2022; Stutterheim et al., 2020; Stutterheim & Gerwien, 2023).

Two key points can be derived from Stutterheim’s and Klein’s research on events for the current context. First, while the research by Stutterheim, Gerwien, and colleagues focuses on event segmentation (cf. Zacks et al., 2007), it suggests that an event representation can encompass multiple event layers. Concerning the internal structure of multilayered event representation, they assume that event layers consist of bundles of non-hierarchical structures, where one or more event layers such as *path* or *manner of motion* are cognitively represented as two sides of the same event, without any specific internal, logical relationship between them. In this dissertation, it will be examined whether this is consistent with ideas found in the philosophical literature on the subject that have been translated into hypotheses about cognitive representation (Goldman, 1970; Löbner, 2021). As will be shown by analyzing selected central examples, multilayered event representations may be explained as hierarchical taxonomies of logical relations. It will then be further explored how the hypothesis of hierarchical taxonomies of events fits with theories of lexical semantics of verbs (Löbner, 2021; Van Valin & LaPolla, 1997) and with theories of Construction Grammar (e.g., Goldberg, 1995). As we will see, the idea of multilayered taxonomies is not completely foreign to syntactic and semantic theory. This is related to the principle that events expressed through language consist of abstract schematic structures instantiated by more specific lexical concepts.

Second, while Klein’s research originally focused on the lexical semantics of (e.g.) event-denoting verbs, his theory allows us to formulate the prediction that the mental representation of these verb meanings (i.e., event structures associated with the verb) can be composed of bundles of temporally complex cognitive structures, and, furthermore, that internal structures can be accessed through language (i.e., through grammatical

operations on the verb). The implications of this hypothesis for speech processing are investigated in the methodological part of the dissertation. By measuring eye-movements during speech comprehension, visual-world experiments are used to investigate how grammatical structure (at the morphological and syntactic level) interacts with event structure. To this end, double-object verbs and sentence structures conveying transfer-of-possession meaning are used. First, the study focuses on how the grammatical aspect marking on double-object verbs modulates temporal features of event representation during online comprehension and can be used to predict event participants and their states (Experiment 1). Second, the experimental study focuses on how the abstract syntactic structure of the double-object construction is used independently of the lexical verb in the integration of event participants that are assumed to be associated with transfer-of-possession meaning (Experiment 2).

Together, the theoretical and experimental study will help to capture and understand new dimensions of the cognitive structures that underlie our ability to mentally represent and process events in the external world. The goal is to provide a basis for a deeper investigation of what we understand when we perceive an action performed in the world, either through language or direct observation. Chapters 2–4 lay the groundwork by examining the metaphysics and cognitive representation of actions and events, introducing the concept of *cascades* to model multilayered event representations, and exploring *Argument-time Structures* and *Intersecting Object Histories* to capture the dynamic, temporally complex nature of events. Drawing from Construction Grammar, these chapters explore how abstract syntactic forms encode such event representations. Chapters 6–8 present empirical investigations testing key theoretical claims. These chapters explore how abstract syntax and grammatical aspect interact to shape sentence interpretation, examining whether sentence forms activate basic event structures and whether aspect is used in real time to perspectivize these. Experiment 1 focuses on how grammatical aspect interacts with event representations during real-time comprehension, showing how aspectual marking highlights specific temporal phases of transfer events. Experiment 2 examines the syntactic encoding of event structures by investigating whether double-object constructions, even with novel verbs, can convey transfer meanings independently of

lexicalized ditransitive verbs. The findings support the hypothesis, grounded in Construction Grammar, that sentence forms encode core event meanings through their structural relationships. Chapters 9 and 10 synthesize the theoretical and empirical contributions while critically evaluating their implications. Chapter 9 integrates the experimental results with the theoretical framework, emphasizing the interplay between temporal complexity in event cognition and linguistic encoding. Chapter 10 focuses on the limitations of the study and addresses challenges to Construction Grammar. First, it highlights the distinction between listener-based and speaker-based approaches to validating Construction Grammar. Second, it discusses its proponents' claims about how constructions emerge as generalizations induced from input. By identifying challenges to Construction Grammar, Chapter 10 outlines directions for future research aimed at refining and extending the framework.

2 The metaphysics of actions and events

2.1 Events and objects

Few would deny that intentions, actions, and changes of physical state are embodied in the entities of the world. When a cook cuts an onion, she can simultaneously have the intention to cook for satisfied guests, to prepare an onion soup, and, even more fine-grained and basic, to operate a knife by controlling her motor skills so that each mentioned and additional conceivable goal is achieved. Each intention, action and resulting change of state of, e.g., an onion being cut, are closely connected to the cook and her instruments. Nevertheless, few would deny that each of these intentions, actions, and changes of state is more than the mere existence of the objects involved (e.g., the people, instruments, and other materials). Intuitively, an onion cannot be simply equated with the cutting of an onion; rather, there is additionally something happening to the object that cannot be explained by its existence alone.

While an object occupies a clearly defined spatial domain of the world and can be clearly separated from other materials, the relation of its existence to the temporal dimension is rather vague. Although an object does not exist outside of time and it is neither infinite nor unchangeable (an onion grows, ripens, and rots), these properties play a minor role for the human consciousness about its existence – it can simply exist in a certain place at a certain time – and it seems secondary next to recognizing its existence in the spatial dimension. However, in the case where an onion is cut, the temporal dimension receives a prominent role. The reason for this is the fact that the transition of the state of the onion from uncut to cut necessarily unfolds over time. In philosophy, such phenomena that *happen* are categorized as events. Events are distinguished from the mere existence of their involved objects by their unfolding over time and associated changes in object properties (Casati & Varzi, 2020; but see Galton & Mizoguchi, 2009).

The relationship between events and the objects, from which they cannot be dissociated, has been under debate. That is, the question of whether events are to be equated with

their constituent material, or whether events constitute an independent ontological category. One position involves the understanding that events are simply unstable objects, by which events and objects are considered as being two aspects of the same ontological entity. In this sense, the only difference between an intact onion and an onion being cut is that the former is more stable and internally coherent, and the latter is more unstable and changes over time (Goodman, 1951; Quine, 1970). Hence, this view includes the idea that events cannot hold an ontological status that is distinct from their constituent material (i.e., their involved objects) (see, e.g., Galton & Mizoguchi, 2009).

Another position draws a line between events and objects and assigns them different ontological statuses (Bennett, 2002). A delimitation of events and their objects can be derived from the following argumentation: First, the event (e.g., cutting an onion) cannot be equated with the object per se (i.e., the onion) because the object also exists outside the time of the event (hence, the onion does not exist only at the time of it being cut). Second, assuming that the event can be equated with the object during the time of the event, challenges arise in determining which of the potentially simultaneous events constituted by the object would equate with the object. When an onion is cut, the onion may be simultaneously moved on the cutting board, it may be warmed up by the hands of the cook, etc., and each such conceivable event may be associated with the same material at the same time. Following Bennett (2002), we will assume for the current context that each of these events can be distinguished from the others, and, moreover, that these events that coincide in time and space, such as ‘heating up an onion’ and ‘cutting an onion’, are to be understood as layers of one “thick” event. Because of this multilayeredness of simultaneously occurring events occupying the same spatio-temporal domain (i.e., instantiating an event token, or, according to Bennett (2002): a “property instance”), the event is for the current purpose assumed to transcend its constituent objects, thereby obtaining an ontologically distinct status.

2.2 Cascades: Multilayered structures in human action

As briefly sketched above, there are a priori arguments for assuming that events are entities that transcend their constituent objects. Such an assumption implies that events have a distinct status in the world and can be regarded as entities independently of objects. Moreover, it follows that several events can be associated with one object. This idea has been suggested by Bennett's (2002) cannonball-example: If one imagines a cannonball flying through the air, one can perceive two events simultaneously associated with the cannonball (i.e., with one object): The cannonball can fly through the air and meanwhile rotate around itself. Together, both events – the movement and the rotation – form a "thick" event according to Bennett.

In this chapter, we will focus on events that instantiate multiple events simultaneously. We will look at ideas that have been introduced in ontology in the domain of human action, and which have been adopted in the description of cognitive structures and verb semantics. The relevant aspect of verb semantics involves a subtype of events involving actions performed by humans. For example, while the sentence *The cook cut the onion* specifies a basic event structure *who did what to whom*, the lexical content of the verb cut in addition specifies the qualities of the action involved in the event, hence, *what* was done (see Konopka, 2018).

In the context of the above definition of events and their involved human actions, a prediction arises that a human action can instantiate several types of actions simultaneously. A person crossing a street can simultaneously move in a goal-directed manner while performing the movement with a specific gait (walking/jogging/running). The fact that the gait of the person can vary suggests that it should be considered as a distinct action independently of the event of the person's goal-directed movement. Goldman's (1970) *A Theory of Human Action* and Löbner's (2020, 2021) modified version on the fundamental structure of human actions serve to shed light on the structuring of such multilayered phenomena. While the explanatory potential of Goldman's theory was first limited to ontology, Löbner adopted principles from Goldman to explain the cognitive representation of the same phenomena. In the following, principles and terminologies are adopted from both Goldman and Löbner.

Goldman's theory and Löbner's extended version are based on the intuitive idea that a concrete human action or *act-token* represents more than one act type at the same time (for an elaboration of the notion of human action and its delimitation to the notion of events, see chapter 2.2). Multiple act types are thus tokenized/instantiated by the activity of an agent occurring in time and space. Bennett (2002) characterizes such instantiations as "property instances", or concrete exemplifications of a universal property that can apply to any relevant entity in the world. Similarly to Bennett, Goldman and Löbner assume that multiple actions can be instantiated by the same entity (i.e., object). In this sense, a human action instantiates a multilayered structure of different act types at once.

Goldman uses so-called act tree diagrams to represent the multilayered nature of complex action structures. A fundamental principle in Goldman's modeling is based on the notion of level-generation. This term captures different logical dependencies between layers (i.e., levels each containing an act-token) that are accomplished by performing an action. Following Löbner (2020, 2021), act trees will from now on be referred to as cascades.

It is important to mention here that the principle of a cascade does not only include sequences of several actions or events, but it also includes the parallelism of several actions that occur simultaneously and are connected to the same object in the sense of Bennett (2002). Cascades in this sense comprise both serial as well as parallel structural relations.

Level-generation between the layers (i.e., act-tokens) A and A' is symbolized by an upward arrow (↥). The most important logical types of dependencies between act-tokens will be illustrated by the example of a few selected actions. First, we are concerned with cascade relations (i.e., level-generation) of the causal and conventional type, which are structured according to the pattern A ↥ A' ('the agent does *c* ↥ the agent causes *e*'). Fundamental relations of this type are irreflexive (if A ↥ A', then A and A' are different), transitive (they form a chain according to the principle that if A ↥ A' and A' ↥ A'', then A ↥ A''), and asymmetrical (if A ↥ A', then not A' ↥ A).

Then, special cases of the so-called simple type, which according to Goldman's and Löbner's analyses exhibit other properties than the causal and conventional types, are illustrated by way of example. The simple type is fundamentally different from the others in that it does not contain causality between levels. In the present analysis, the simple type is presented as a problematic case for the consistency of cascade theory. To meet this challenge to cascade theory, a further counterfactual type will be discussed for the present analysis in chapter 2.3. The counterfactual type of level-generation is intended to capture logical dependencies in any cascade according to similar principles as the causal and conventional types. However, it will be proposed as a novel addition to Goldman's and Löbner's ideas on the basis on Lewis' (1973) counterfactual analysis.

2.2.1 Causal level-generation

In (1), we see examples of actions that typically take place during a chess game (Goldman, 1970: p. 21).

(1) A' x moves a chess piece
↥
A x moves her hand

The minimal action cascade shown in (1) reads from bottom to top. It starts on the lower level (A) with the intentional action of an agent (x). Here, the question of whether A constitutes the most basic action is not addressed; the action specified here is at least approximately basic and serves as an intuitive anchoring of a conceivable perceptual experience. Moreover, there is no upper limit to how many further or higher levels can be generated by A. The present example serves solely as a representation of a minimal cascade showing that as one action is performed, other actions are typically performed as well.

The plausibility of the ontological reality in the presented cascade (1) is straightforward to intuitively test: A chess player must move her hand in order to move a chess

piece. There is a causal dependency between the actions: The chess piece is moved as a consequence of an action of an agent. Both actions take place strictly in parallel in time and space, occupying the same spatio-temporal zone (cf. Bennett, 2002). Goldman's definition of causal level generation basically states that when an agent does A, the agent also does A', with A' being generated by A.

Consider in (2) another example by Goldman of causal level generation that could potentially stand as an extension of (1), i.e., as higher levels of the same cascade.

(2) A' x gives her opponent a heart attack
↥
A x checkmates her opponent

One intuitively sees in (2) that the relation between the levels A and A' is causal: By the fact that x puts her opponent into checkmate, x also causes the heart attack of the opponent. In contrast to (1), A' in (2) does not occur strictly at the same time as A, but presumably somewhat delayed. Goldman points out, however, that although A and A' do not happen parallel in time, A and A' still take place at one time. This is because an agent does A' by doing A. Therefore, the time of the action is (in a certain sense) defined by the total time period that is completed at each cascade level. Example (1) and (2) show that multiple act-tokens generated by the same action can take place independently of their position in time and space. This rather abstract definition of temporal relations – or of what marks the beginning and the end of an action – is important because it allows us to see in a consistent way which aspects are contained in an action, hence, which layers of an action can potentially be identified. This becomes particularly relevant for the discussion of how language interacts with cascades (see chapter 4). For the moment, we will remain with the ontological perspective and consider further logical types of level-generation.

2.2.2 Conventional level-generation

In (3), we see a section between the cascades shown in (1) and (2).

(3) A' x checkmates her opponent
↥
A x moves a chess piece

The causal level-generation that exists in (3) between A and A' succeeds only under specific conventional circumstances. Precondition for this level-generation is first of all that each of the involved people follow the rules of the chess game. Goldman speaks in this context about conventional level-generation as another type of level-generation next to the causal type (see the previous chapter).

This means that the logical type of level-generation varies from level to level of the same cascade. In (1), we see a cascade which contains a causal relation: The movement of a hand causes the movement of a chess piece. For the relation between the movement of a chess piece and the action of putting someone into checkmate, there is a conventional circumstance which is determined by the rules (i.e., conventions) of a chess game.

If we look at another example in (4), we will also notice a social convention.

(4) A' x causes her child to not possess her or his computer
↥
A x forbids her child to play on the computer

In (4), the utterance of a prohibition A implies the simultaneous execution of a further action A' which in turn depends on A. The prohibition thereby causes someone to undergo a loss of possession. A and A' do not take place strictly at the same time, because a prohibition (A) must first be uttered before the further successive action (A') can take place. However, by doing A, an agent also does A' – under the condition that social conventions are followed. That is, the implicature of the conventional (and at the same time also causal) level-generation A ↥ A' of example (4) consists in the fact that the concrete

tokenization of event *e* (A) under specific social conditions implies that event *c* (A') also takes place.

2.2.3 Simple level-generation

Goldman understands relations between actions primarily as causal and conventional dependencies. However, as we have seen in chapter 2.1, relations of the causal and conventional type do not describe every type of multilayered action. In fact, there are also relations that cannot be easily classified as hierarchical. If one imagines a cannonball flying through the air while rotating around itself (cf. Bennett, 2002), it becomes clear that both events – the movement and the rotation of the cannonball – are temporally parallel and occupy the same spatio-temporal zone, i.e., are associated with the same object at the same time. Although the inertia of both physical phenomena is caused by the same causal force (the firing of a cannon caused at a lower level of action by the intentional action of an agent), there is no causal relationship between the trajectory and the rotation of the cannonball: It does not fly because it rotates, neither does it rotate because it flies (but because it was fired from a cannon).

Goldman subsumes parallel, non-hierarchical phenomena of this kind by another type of structural relation which he refers to as *simple level-generation*. Goldman notes that actions such as 'hitting the tallest man in the room' and 'hitting the wealthiest man in the room', where the tallest man and the wealthiest man refer to the same person, does not exhibit level-generation. Rather, the symmetry of such actions suggests that they must be represented at the same level of the cascade:

> "[…] it is natural to regard these two acts as being on the same level, and to represent them by circles connected with a horizontal line. Moreover, we can stipulate that if A and A' are same-level act-tokens, then any act which generates A also generates A'; and any act which is generated by A is also generated by A'." (1970: S. 31).

Goldmans does not go any further with modeling simple level-generation, but only outlines that act-tokens of the same cascade level would have to be represented in parallel. However, he does not elaborate on how *level-generation* can occur between levels of a

cascade that, since being on the same level, cannot generate each other. In other words, it remains unclear how, according to Goldman's formalization, A and A' must both be generated symmetrically, without A' being generated by A. Nevertheless, it will become clearer in 2.3 that this logical relation between levels of action, originally identified by Goldman as a same-level relation, is relevant to the representation of multilayered structures in ontology.

2.3 Counterfactual dependencies in multilayered structures

In this chapter, we focus on a type of level-generation that is not captured by causal or conventional dependencies between act-tokens (i.e., levels of a cascade; see chapter 2.2). It involves the representation of act-tokens that occupy the same spatio-temporal zone (cf. Bennett, 2002). Goldman (1970) has identified such phenomena using the example of the parallelism of the act-tokens (a) 'x hits the tallest man in the room' and (b) 'x hits the richest man in the room', where both 'the tallest man' and 'the richest man' refer to the same person. Neither of the act-tokens generates the other, which is why, according to Goldman, both should be represented at the same cascade level. However, Goldman does not give an explicit proposal for the representation of so-called same-level representations in the context of a cascade. In Löbner's (2021) modified version of Goldman's theory, this logical type of level-generation is acknowledged but referred to as "a constellation of facts" (p. 270); however, it is not further discussed either.

If we take a closer look at the example identified by Goldman, some issues arise from categorizing this example as a multilayered cascade. In particular, in both Goldman (1970) and Löbner's modified version (2020; 2021), it remains unclear to what extent the principle that an act-token constitutes a cascade level and generates further actions irreflexively and transitively can be consistent with the idea of the same-level representations that are implied by level-generation of the simple type.

In the following, for the sake of simplicity and consistency of a cascade-theoretic approach, the concept of same-level generation (level generation of the simple type) will be abandoned. Instead, it will be attempted to understand any instance of multilayeredness

in human action as a strict hierarchy, i.e., to structure cascades so that an act-token can only generate one other act-token.[1] As Löbner (2021: p. 270) suggests, (a) and (b) do not refer to a set of same-level act-tokens, but rather to "a constellation of facts". Here, it is assumed that both (a) and (b) describe only one and the same tokenization of the act-type 'x hits y'. This is because of the fact that the entity y can be referred to by any of its properties is a phenomenon of reference, not of level-generation. Here, (a) and (b) are not considered as different act-tokens generated by an agent (contra Goldman), but one single act-token that happens to be described in different ways.

[1] As noted by Goldman and Löbner (2021, p. 270), is it important to mention that causal level-generation can branch into multiple parallel act-tokens that are not hierarchically categorizable: An act-token declining the nomination for vice-president can generate the two branching act-tokens (a) ↥ 'breaking a long-standing tradition' (an example of simple level-generation because (a) arises from a co-existing circumstance (i.e., tradition)) and (b) ↥ 'disappoints his followers' (causal level-generation). In the current context, the dependencies between tokens of the same action are at question; this dependency is not present between (a) and (b).

In the following, a strictly hierarchical approach to cascade theory we be approached by assuming a novel subtype of causal level-generation, namely *counterfactual level-generation*.

2.3.1 Causality and counterfactuality

The basic idea behind counterfactual analyses of causal dependencies can be traced back to David Hume's *An Enquiry concerning Human Understanding* (1748). In his work, Hume discovered a causal dependence between cause *c* and effect *e* in which *c* does not cause *e*, but rather *c* presupposes *e*. The principle is that if the first entity had not existed, the second would not have existed. Lewis (1973) defines Hume's principle as follows:

> "We think of a cause as something that makes a difference, and the difference it makes must be a difference from what would have happened without it. Had it been absent, its effects – some of them, at least, and usually all – would have been absent as well" (p. 557).

According to Lewis' analysis, the term counterfactuality describes a type of causal dependency:

> "Where c and e are two distinct possible events, e causally depends on c if and only if, if c were to occur e would occur; and if c were not to occur e would not occur." (p. 562; quote from Menzies & Beebee (2009)).

According to Lewis' definition, causal dependence thus comprises two types of relations. On the one hand, there can be causality between effect *c* and cause *e*, which can be expressed by the following prediction: If event *c* happened, event *e* would also happen. On the other hand, this implies counterfactuality, which Lewis likewise subsumes as a type of causal dependence. This relation predicts that event *e* would not happen if event *c* would not happen either. Thus, causality is not necessarily present in dependencies between different events *c* and *e*, but only counterfactuality. If we recall the types described above, examples of counterfactual dependence without causality become apparent in level-generation of the simple type. Hence, the relation between act-tokens A and A' in

same-level representations (such as simple level-generation) does not exhibit causality but only counterfactuality between events *c* and *e*.

In the current context, it is assumed that each type of level-generation can be characterized as irreflexive, transitive, asymmetric, and counterfactual. By identifying counterfactual dependencies, we thus avoid having to categorize some cases of level-generation as causal and asymmetric, and others deviating from them as non-causal and symmetric same-level representations. This assumption allows us to classify each cascade as a counterfactual hierarchy according to the principle that one act-token generates another act-token at a higher level. Causality between levels (or actions or events) may be present, which can be predicted for causal and conventional level-generation by applying Lewis' definition.

In summary, according to the principles described above, level-generation of the simple type does not have to be assigned a distinct, idiosyncratic status from the causal and conventional type. It will be shown in chapter 2.3.2 that relations between act-tokens occupying the same spatio-temporal zone (cf. Bennett, 2002) exhibit counterfactual dependencies without causality between A and A'. Recognizing counterfactual relations between tokens of the same action (or layers of an event) therefore allows us to understand each type of level-generation as a hierarchical taxonomy. This assumption might address challenges for cascade theory that arise with respect to the simplicity and consistency of a cascade model, which will be demonstrated in chapter 3 by considering a cognitive perspective on cascades.

2.3.2 Level-generation as a counterfactual dependency

As described above, cascade theory faces the challenge of trying to structure hierarchies of act-tokens on the one hand, while on the other hand for some types of level-generation having to accept that multiple, non-hierarchical act-tokens must be represented at the same level. This chapter aims to support the assumption that each act-token is connected to a hierarchically subordinate act-token by counterfactual dependency. The goal here is to challenge the claim by Goldman (1970) and Löbner (2020; 2021) that some act-tokens

must be represented at the same level in a hierarchy. It will be attempted to solve the problem of modeling same-level act-tokens in the context of a strictly hierarchical cascade theory. Starting with an example from Bennett (2002), the level-generation of the simple type will be analyzed as a hierarchical structure of counterfactual dependency.

Let us examine in more detail which type of logical relation represents level-generation of the simple type. If we recall the above-mentioned example of the cannonball flying through the air and rotating around itself at the same time, we can assume that, following Goldman and Löbner, levels of such multilayered structures in principle cannot be divided hierarchically. Rather, they would be classified as a case of simple level-generation: A causal force causes the cannonball to move through the air and rotate around itself, with both the movement and the rotation being generated as two different act-tokens on the same level. Both act-tokens would have to be represented on the same level according to principles of simple level-generation, because neither the movement nor the rotation causally generates the other.

In (5) we see a minimal section of a more complex cascade that classifies the phenomenon of a moving and rotating object as a hierarchy of act-tokens.

(5) A' x causes the rotation of y
↥
A x causes the motion of y

In (5), the act-token A' is generated by A. Paraphrased, it means that x causes the motion of y, and in doing so, x causes the rotation of y. A' is generated by A because the level-generation 'x causes the rotation of y' ↥ 'x causes the motion of y' seems implausible. In (5), there is no causality between A and A', but rather, this relation is counterfactual: without the act-token A, A' would not occur, i.e., the rotation of y could not occur without motion of y under the conditions that the physical properties of the world are predictable.

As can be induced by experience with the physical laws of the world, a moving and rotating object needs a causal force to initiate both. The causality of the cascade exists between A and the two levels A' and A'', which we see in (6):

(6) A’’x causes the rotation of y

↥

A’ x causes the motion of y

↥

A x causes the firing of y

To what extent the phenomenon of a moving and rotating cannonball is conceivable in independence of the causal force that caused both events associated with the cannonball is not considered here. Hence, for the time being, we do not consider the cognitive perspective on such a cascade.

In summary, (5) and (6) serve as examples to identify different causal dependencies: On the one hand, it is implausible to assign physical properties (motion and rotation) to y without being able to trace back a causal force x. On the other hand, although motion cannot be a cause of the rotation of y, the rotation of y would not occur without motion. Thus, the causal dependence between the motion and rotation of y must be counterfactual without causality.

In (8), we see another example of a similar phenomenon, where a leaf falls and rotates at the same time. In this event, it is assumed that a natural force can be represented analogously to human agency (this will be justified in more detail when considering the cognitive perspective on cascades in chapter 3). To this end, the simplified example is to be understood under the specific conditions, e.g., that air resistance is present.

(7) A’’gravity causes the rotation of the leaf

↥

A’ gravity causes the fall of the leaf

↥

A gravity pulls a leaf from the tree

In (7), gravity could also be replaced by the blowing of the wind as causal force behind the motion and rotation from the leaf. In this assumed ontological phenomenon, it only plays a role that the generation of A' and A'' must be caused by a natural force. As in example (6), the rotation is counterfactually dependent on the motion.

In the domain of human action, examples of counterfactual dependencies are frequent and fundamental. (8) shows a walk/jog/run cascade.

(8) A' x walks/jogs/runs
↥
A x moves

Example (8) describes the motion- and manner-levels involved in a motion event (cf. Talmy, 2000). Here, A' refers to a specific way in which a person moves. Importantly, A' would not take place without A taking place: x moves in a specific manner by moving in the first place. Analogous to examples (6) and (7), in which a specific manner of movement is also accomplished by an entity moving, the counterfactual dependency only goes in one direction between A and A'. One might object that by walking/jogging/running, it causes one to move along a spatial path. This is illustrated in (9). However, this goal-directed movement is different from the fact that manner of movement presupposes movement. Hence, a path component must be additionally conceptualized (cf. Talmy (2000) who defines manner as a co-event to motion).

(9) A' x moves in the direction of the Eiffel Tower
↥
A x walks/jogs/runs

(9) represents the fact that one cannot move in a goal-oriented direction without moving in a certain manner. Hence, one reaches a goal or moves along a path by walking/jogging/running. Lower in the cascade, one might imagine the subordinate levels 'x has the

intention to achieve a goal' ↥ 'x has the intention to control her body motor activity to achieve the goal' ↥ 'x moves'. Whether there is causality between A' and A in (9) is not relevant here[2].

Finally, let us direct attention to an example given by Goldman that illustrates level-generation of the simple type, where neither 'x hits the richest man in the room' could level-generate 'x hits the biggest man in the room', nor the other way around. As Goldman notes, this example cannot be structured hierarchically. For this reason, he assumes that A and A' are to be represented symmetrically at the same level in the cascade. At first glance, Goldman's prediction does not run into problems in examining if there is counterfactuality or causality. It is argued here that Goldman does not encounter any challenges in this since none of the two causal dependencies can be identified between (a) and (b). It is assumed here that logical dependencies between (a) and (b) are absent for the

2 One might object that the intention to move causes one to reach a goal, e.g., the Eiffel Tower. However, this is only true if the intention of x is to move to the goal. In this case, the fact that the Eiffel Tower could be reached would be understood as a coincidental circumstance.

reason that (a) and (b) describe only one and the same act-token. The reason for this is that unlike in example (9), specifying properties of the target of the hitting-action ('the richest man' and 'the tallest man') does not equal specifying properties of the action itself.

Thus, Goldman's example is assumed here not to represent parallel representations of multiple act-tokens, and thereby not to represent properties of the act itself, but only referential properties of the goal of the act. Although Löbner (2021) adopts Goldman's non-hierarchical and symmetrical premise for simple level-generation, Löbner describes Goldman's example as a constellation of facts. However, it can be argued in accordance with Bennett (2002) that, unlike act-tokens, facts of this kind are not spatio-temporally bounded constructs. The facts that someone is the richest man on the one hand and the tallest man on the other (in addition, the man under discussion possesses a potentially unlimited number of other conceivable properties) are true for him even before and after the action in which someone hits him.

One could object that facts in principle can also be actions and events. E.g., it is a fact that the event 'the Battle of Waterloo' took place, and by the fact that it did, this factual event is an act that took place in time and space (i.e., an act-token), which involved the generation of several act-tokens at the same time. However, the facts identified by Löbner are not events nor actions, but mere properties of an entity that can be understood independently of any action performed by or on this entity.

3 Actions and events in cognition

So far, we have looked at ideas about the ontology of actions and events, which include the observation that events and their involved actions taking place in the world may be intrinsically multilayered in nature. However, the aspects of ideas presented do not make any claims about how we understand and mentally represent actions and events. In the following subchapters, we will look at insights from Stutterheim, Gerwien, and colleagues about multilayered representation, which will then be compared to the theory of multilayered cascades presented in the previous chapter. The chapter will serve to provide a basis for defining cognitive structures that go beyond the commonly assumed format of semantic structures or event structures.

3.1 Multilayered cognitive structures

The assignment of events as entities of a metaphysical distinct status requires the assumption, on the one hand, that events transcend their constituent objects and, on the other hand, an explanation of how several events instantiated by the same object are logically connected. If we turn our attention to the cognition of such phenomena, it follows from the fact that we are able to imagine and perceptually experience events that events of the external world must have some kind of mental correlate. This assumption is what connects the various cognitive science disciplines of philosophy, psychology, and linguistics, which from that point on represent rather diffuse epistemological interests (cf. Fuchs, 2017). Those theories that seek to explain the mental representation of outer-worldly events, however, face similar explanatory challenges as those that pursue their ontological nature: They need to explain structural relationships between events, between internal event components such as actions, as well as between these and their linguistic expression.

Studies concerned with the visual perception and linguistic expression of events have made an observation similar to that of Bennett (2002) (see chapter 2.1). Based on the

notion that events in the world can coincide in time and space, Stutterheim, Gerwien, and colleagues have hypothesized that events are composed of layers (Gerwien & Stutterheim, 2022; Gerwien & von Stutterheim, 2022; Lambert et al., 2022; Stutterheim et al., 2020; Stutterheim & Gerwien, 2023). Through experimental studies, this line of research has investigated how the focus of attention on a particular event layer determines how events are conceptualized and, consequently, divided into units. This idea can be summarized with the following observation: When perceiving someone running across a train station and boarding a train, we can either perceive the event of someone moving along a path to a goal, which involves two separate events ('crossing the station' and 'boarding the train'), or we can perceive a single, more specific event of someone moving in a particular way (running). Conceptualization such an event can therefore be based on the layer which receives the initial attention. Moreover, central to this line of research is the observation that languages differ as to which of these layers is typically expressed by the verb. In French and Tunisian Arabic, the path of a moving entity is typically expressed by the verb. In German and English, on the other hand, the verb typically expresses the manner of motion of that entity (walk/jog/run), while the path can be encoded within the same proposition by adverbial modification of the verb (e.g., *across the station into the train*). Typological variations thus lead to differences in how many layers of an event the speaker of a specific language can express through one proposition. Experimental evidence shows that when speakers from the different typological categories describe a motion event (depicted in a dynamic, visual scene), areas of the scene that can be associated with either the manner of motion of an entity or with the path of motion (i.e., the goal of motion) attract eye-movements in accordance with how the scenes are linguistically encoded. Furthermore, Gerwien and Stutterheim (2018) found evidence that even in non-linguistic contexts (without any description tasks), events are segmented according to principles of their linguistic expression.

Although this line of research focuses especially on how events are segmented (cf. Zacks et al., 2007), it can be used to generate hypotheses for other cognitive processes. Arguably, this line of research predicts that multiple event layers can be instantiated within one event representation in working memory. Regarding the architecture of this

aspect of mental representation, Stutterheim, Gerwien, and colleagues claim that event layers are structured non-hierarchically, i.e., non-taxonomically (see Gerwien, Marberg & Nicolaisen, *in press*). Hence, their approach predicts that events consist of unstructured bundles of event layers, each of which can be brought into attention. In this way, the authors' assumptions differ from the theory of human action as proposed by Goldman (1970) and later applied to a cognitive context by Löbner (2020; 2021).

As we saw in chapter 2, Goldman and Löbner capture events as hierarchical taxonomic structures from both a metaphysical and a cognitive perspective. For Goldman and Löbner, the instantiation of an action generates a cascade of associated layers of actions that are carried out in parallel. The structure of a cascade potentially includes causal dependencies between layers, by which it is described how actions bring about state changes of affected entities, and non-causal dependencies, by which counterfactual dependencies between multiple co-temporal state changes of the same object are described. Any state change induced by an action is understood here synonymously with the term event.

One could argue that an equation between, on the one hand, events as entities defined solely by changes of objects (see chapter 2.1) and, on the other hand, Goldman's theory focusing on the nature of human actions, would limit the explanatory potential of a cascade-theoretic approach exclusively to events that involve human actions. For example, since a cannonball flying through the air or a leaf falling from a tree qualify as events (since state changes of the cannonball and the leaf are present), but not human actions, it could suggest that a cascade-theorical approach does not include these events. However, a closer look shows that the action of a human can be thought of and de facto traced back as the cause of the movement of a cannonball. In the case of a falling leaf, although no human cause can be traced back, it becomes apparent that the action of a breeze or the fragility of a twig that must yield to gravity is nevertheless comparable to the causality of human action. Since we currently remain agnostic towards the ontological perspective and want to make statements only about the psychologization of cascades, this comparison can be justified by the following assumption: The cognition of causal potentials of the world includes not only actions of living beings (from the level of microbiology) but also forces of nature (from the level of particles). This parallelism in the cognitive

processing of both phenomena is evident from semantic restrictions of causative arguments, that, in addition to humans, allow natural forces (*The rain swept the ring into the gutter*) but not instruments (**The broom swept the ring into the gutter*) (Goldberg, 1995: p. 165). Following this idea, it is assumed here that an analogy can be drawn between the causal potential of a natural force (e.g. gravity, wind, etc.) and the agency of a human being (or the causal potential that can be attributed to human actions, such as a metonymical understanding of *war* in *The war brought misery to everybody*). For the above-mentioned example with a falling leaf, this allows us to classify a natural force as a lower cascade level, in that gravity causes the motion of the leaf.

According to Löbner's interpretation, principles of Goldman's theory can be applied to cognition. The main focus is not on specific contents of mental representations, but on the structuring of mental representations resulting from the processing of perceptual experiences of the inner or the outer world. In other words, the explanatory potential of such a perspective on mental representation does not lie in defining the contents of generalized knowledge in long-term memory, but rather in adopting a processual perspective of how the temporary retrieval and processing of this knowledge in short-term memory works. Hence, for Goldman and Löbner, a cascade represents not abstract types of actions, but act-tokens that can be though to take place in time and space.

To describe mental formats of temporary representation, notions such as *mental models* (Johnson-Laird, 1983) and *situation models* (Van Dijk & Kintsch, 1983) have been used in the literature. Mental models are regarded as multi-modal representations in which information from perceptual input is combined with knowledge from long-term memory (e.g., schemata), or by which analogous processes of the internal imagination take place. Mental models capture multimodal simulations of complex relationships in the world, including factual knowledge (e.g., the knowledge that the tallest man in the room is also the richest, as per Goldman's (1970) and Löbner's (2021) example), visual properties such as colors and shapes, and the format of event representations relevant to the current context. Mental models allow us to select events piece by piece and represent or adapt them for descriptions. The idea of a mental model therefore implies that they are composed of multiple event representations (see Gerwien, Marberg & Nicolaisen, *in press*).

It is assumed here that the hypothesis that working memory simultaneously represents multiple events is compatible with a cascade-theoretic view of multi-layer instantiations of act-tokens. Each act-token can thereby be captured as a proposition. It is thus assumed that cascade theory can be understood as a cognitive theory about the logical dependencies between a mentally highlighted propositional event structure and other conceivable event structures of the same cascade. From this point of view, cascade theory is useful as an explanation for the logical dependencies between conceptualized event units between which we can toggle our attention. If it is assumed that a cascade level represents an event, it can explain that when we imagine (i.e., mentally instantiate) a moving and rotating cannonball, not both the movement and the rotation must necessarily be represented at once, but one of the events instantiated by the cannonball may alone receive full attention. The idea can be captured by another example involving human action: If we imagine the event of a child folding a paper airplane, we can switch our inner attention between the parallel events (i) the manner by which the child is acting on a piece of paper (folding); (ii) the changing state of the paper (changing from flat to folded); and (iii) the child's intention to achieve a goal (making a paper airplane) – or we can represent (i), (ii), and (iii) all at once, i.e. all cascade levels can be mentally instantiated.

More precise internal structurings of instantiated event representations do not fall within the focus of cascade theory. In the next chapter, the internal properties of instantiated event representations will be examined from two perspectives, one considering logical-temporal structures (Klein, 1999; 2002) of events and the other considering the role of event-constituting objects (Altmann & Ekves, 2019).

3.1.1 The mental representation of events and objects

In research on perception, the notion *events* refers to mental constructs that allow us to understand and interact with the world. The formation of mental constructions of this type requires that a dynamic perceptual input is segmented into units with beginnings and ends (Zacks et al., 2007; Zacks & Tversky, 2001). This is essentially accomplished by recognizing the types of entities outlined earlier in this dissertation: On the one hand, objects

(e.g., humans and objects), and, on the other hand, the more abstract schematic content of events that defines logical and temporal configurations of the involved entities.

A widely used approach to describing the contents of event representations goes back to Fillmore's (1968) case grammar theory. According to Fillmore, event representations are composed of a finite set of abstract concepts:

> "The case notions comprise a set of universal, presumably innate, concepts which identify certain types of judgments human beings are capable of making about the events that are going on around them, judgments about such matters as who did it, who it happened to, and what got changed."

Fillmore's theory, the core of which is evident in the quote given above, was intended to capture potentially innate generalizations of event entities (i.e., event participants). According to Fillmore's approach, experiences with basic actions of the external world are generalized as conceptual, prototypical schemata in the internal world, which in our terminology are referred to as event representations. Event representations are stored as units of meaningful structures (or *frames*) and allow the embedding of prototypical event participants. In Fillmore's terms, the latter phenomena are called cases; in later research that has adopted the principles of Fillmore's case theory to a lesser or greater extent, they are referred to as semantic or thematic roles. An example of a generalized schema is the action denoted by the double-object construction (DO) with the verb *hand* as in *Laura handed her mother the letter*. The referent *Laura* fills the thematic role of the agent (the entity that performs an action), *her mother* the role of the recipient (the entity that obtains possession) and *the letter* the role of the theme (the entity that undertakes a spatial change of position by changing possession). Each of these thematic roles are understood relative to the conceptual schema and fill in gaps in the schema.

Dowty (1989) points out that research following Fillmore's idea has divided into two different theoretical directions. On the one hand, the notion of thematic roles is understood as configurations of abstract roles. Here, the handing-schema instantiates a more abstract schema that can be composed of an agent, recipient, and theme. This abstract schema thus allows for multiple potential instantiations, e.g., also giving- and sending-events, that are understood relative to the same schematic meaning, in this case an abstract

transfer-of-possession meaning. On the other hand, Fillmore's notion is understood as schemata specifying individual action types. From this point of view, the handing-schema constitutes an action type in itself, with which individual thematic roles are associated (such as 'the one handing something over to someone', 'the one who gets something handing over') (see Gerwien, Marberg & Nicolaisen, *in press*; see also Frankland & Greene (2019)).

For the principles of cascade theory, this distinction does not play a role in the first line: each cascade level constitutes an act-token, i.e., an instantiated event representation. Thus, the idea of a cascade as a processual perspective on event representation in working memory makes no claims about the representation of action types in long-term memory. The perspective on event representations as logical configurations of generalized thematic roles such as agent, theme, patient, recipient, and instrument thus provide an analytic approach to approximate complex cognitive phenomena.

Other theoretical and experimental approaches have additionally focused on the temporal dimension of events (Altmann & Ekves, 2019; Hindy & Altmann, 2012; Misersky, Slivac, Haagort & Flecken, 2021; Klein, 1994; Klein, 1999; Klein, 2002). This research is based, among other things, on the idea that perceptual experiences of events unfold over time, with different event phases being linked together to form a coherence between the 'before' and the 'after' the state changes of the objects involved. Different directions of this line of research are connected by the assumption that conceptual contents of event representations are distributed over different temporal event phases. This is evident in simple events (or the minimal instantiation of an event type) denoted by a DO construction such as *Laura handed the letter to her mother*. The minimal representation of this event, which can be defined by a transfer-of-possession schema with co-located entities (see Beavers & Koontz-Garboden, 2020; Rhode, Kehler & Elman, 2006), is constituted by an initial phase in which the mother has not yet received the letter, and then a final phase in which the transfer is complete, and the mother possesses the letter. Such a temporal component (i.e., the logico-temporal progression of the event), is not captured by Fillmore's case theory.

In the following, two approaches are presented that examine the temporal component of event representation. First, we are concerned with a theory by Klein (1999; 2002; 2010) on so-called *argument-time structures*, which was originally formulated for the analysis of the lexical semantics of verbs but is here applied to the description of cognitive representations. Second, we are concerned with a theory by Altmann and Ekves (2019), who, like Klein, aim to explain basic building blocks of event representations not as thematic roles, but primarily as objects whose state changes over time. The goal is to form a synthesis of both approaches in which events can be captured in greater detail with respect to their temporal and logical structures as well as their linguistic encoding.

3.1.2 Logical and temporal structures in events

Regardless of whether we approach the study of events either as unstable objects or as distinct metaphysical categories, both approaches are defined by the inclusion of a temporal component. Approaches to describing the cognitive representation of phenomena of this type must therefore specify either the instability of a changing object, or a more abstract, schematic representation of logico-temporal relations between objects and their states, or a combination of both perspectives. What ties both perspectives together is the assumption that events are defined by properties of one or more objects changing over time. Hence, in the context of cognition, both ontological perspectives amount to a similar understanding of events. However, as we will see, there are strong reasons to assume that in cognition, events cannot straightforwardly be equated with representations of their constituent objects, but rather, that events can be viewed as abstract schemata. To explore this idea, basic principles of Klein's (1999; 2002) Argument-time Structure Theory (ATS) are presented below.

ATS attempts to explain what constitutes the minimal representation of the meaning of verbs that refer to events in the world. We will assume that the verb semantics specifies properties of event representations that are relevant for linguistic encoding. In chapter 4.4, we will consider some predictions of Klein's theory about the mental processes involved in the linguistic encoding of event representations.

In their essence, argument-time structures are specifications of structural and descriptive contents of a verb. We take these specifications to be informative about what constitutes the minimal content of event representations. On the one hand, they specify causal and counterfactual relations that must be minimally present between event entities and their states in order to make judgments about who does something, to whom it happens, and about what pre- and post-states of changed entities are involved (see also Primus, 2012). On the other hand, they include a descriptive component that specifies qualitative properties, adding conceptual contents to schematic event structures. In (1), an abstract transfer schema is identified.

(1) Laura handed/gave/sent her mother a letter.
ACTIVE <X, t1> & NOT-HAVE <Z, Y, t2> & HAVE <Z, Y, t3>[3]

[3] This representation is taken from Klein's (1999: p. 13) analysis of the lexical content of *schenk-*.

The argument-time structure in (1) reads as follows: The descriptive conceptual property 'ACTIVE' applies to argument X (*Laura*) at time interval t1; 'NOT-HAVE' describes a non-possessive relation between argument Z (*her mother*) and Y (*a letter*) at time interval t2; 'HAVE' describes a possessive relation between Z and Y at t3. In addition to representation (1), this argument-time structure specifies a temporal configuration in which time interval t1 overlaps with t2, and t3 follows t2. A logical configuration specifies a causal dependency between t1 and t3 on the one hand, and a counterfactual dependency between t3 and t2 on the other. Klein explicitly summarizes each of the types of dependency under Lewis' (1973) definition of causal dependency (see chapter 2.3.1). Thus, according to Klein, causal dependencies capture counterfactual relations between pre- and post-states of an object. In (1), t3 counterfactually relates to t2, since argument Z could not obtain possession of Y, if Z prior to this did not have possession of Y.

Thus, (1) describes a causative action type in which – captured as thematic roles – an agent causes the state of a recipient and a theme to change over time. It is important to note that although a change of state of the agent is ontologically conceivable as part of a richer and more complex cognitive representation, the minimal structure of this event merely requires the representation of an agent up to the point where the states of a recipient and theme have changed. The state of the agent at a subsequent time interval, in which the agent has given up possession complementary to the recipient obtaining possession, is not a prerequisite for understanding the causative action.

It is not clear from Klein's analysis, however, to what extent each argument-time structure captures an event representation that goes beyond individual verbs, described above as 'abstract' action types (Dowty, 1989). On the one hand, this seems to be the case in (1), since this argument-time structure is applicable across verbs (*hand*, *give*, *send*) and therefore represents a more abstract action (or event) type than that specified by the particular verb implemented. On the other hand, however, for transitive verbs (i.e., events with a causal relation between two entities), an 'individual' action type (cf. Dowty, 1989) would presumably have to be assumed for each representation due to a greater variability of meanings: In Laura opened the door, the descriptive properties for the door must first be 'not open' and then 'open'; in Laura broke the glass, for the glass, they must be 'not

broken' and then 'broken'. Arguably, each of these descriptive properties can unlike those presented in (1) only apply to these verbs. However, at the same time, it is possible that to understand both events, it requires recognizing a causative action type at a more abstract level, comparable to the more abstract action type (1).

Nevertheless, it is clear that Klein assumes that event schemata can be abstracted from their constituent objects. This can be illustrated by argument-time structure analysis of events that do not comprise state changing entities. Such a phenomenon can be seen by considering the difference between event (4) and (5) (see Klein, 2002: p. 13):

(2) The traffic light turned red.
NOT-RED <X, t1> & RED <X, t2>

(3) The traffic light remained red.
RED <X, t1> & RED <X, t2>

Whereas (2) can be clearly categorized as an event since it contains a change of state, (3) exhibits no change of state of the object involved. In contrary, it describes a situation in which a traffic light remains in the same state of being red. However, it is precisely the absence of a change of state that defines the representation of remain as an event, because the argument-time structure of remain specifies not only a state (for which *be* would be used, as in *The traffic light was red*), but additionally a second state at a subsequent time interval (t2) in which no change has taken place. To recognize the event described in (3) as such, the object state must therefore be compared between two time intervals and recognized as identical against expectation.

In summary, Klein's theory serves as a representation of the logico-temporal structures of event types, each of which can be instantiated as an act-token in terms of cascade theory. Where the explanatory potential of cascade theory does not extend beyond how multilayered events are structured, the argument-time structure analysis aims at examining the schematic representations from an event-internal point of view. In addition, the explanatory potential of Klein's theory goes beyond the view of considering internal

event representations as constellations of thematic roles. By identifying the conceptual components of an event representation as temporally constrained, taking a logico-temporal perspective on event representations allows us to represent causal dependencies and object state changes in a more fine-grained and explicit way. In a transfer-of-possession event, it appears from an argument-time structure analysis that the activity of an agentive entity occurs only in an initial time interval, whereas the properties that are defining for the recipient rather apply to a later time interval for this entity, i.e., the recipient only counts as such by being a final possessor. Moreover, Klein's theory captures and explains a broader variation of events within the same framework, ranging from events that either do or do not involve changes in state of the involved objects (cf. example (3)). This is achieved by assuming that understanding events implies the retrieval of abstract schemata that specify time intervals as well as properties that apply to these time intervals.

3.1.3 Event representations as object representations

Similar to Klein's (1999; 2002) Argument-time Structure Theory (ATS), Altmann and Ekves' (2019) Intersecting Object Histories (IOH) (see also Hindy et al. (2012) for a preliminary version) approach involves seeing event representations as bundles of objects that change over time. Here, IOH and ATS are seen as analogous approaches to a view of events that takes a temporal component with pre- and post-states of event entities into consideration. Unlike Klein's lexical-semantic approach, which in itself makes no claims about cognitive representation, IOH takes a neurocognitive perspective and examines which perceptual properties are relevant to the processing of event representations. Whereas both theories view event representations as specifications of how properties of the involved objects evolve over time, IOH is primarily concerned with how different perceived object states are integrated in the perception of events. In contrast to Klein's theory, IOH does not specify representations of the logically minimal object states that constitute an event, but it presents principles of how the entire course of multiple perceived changes leaves a trajectory of object states in time and space (or the "history" of

an object). It is further assumed that event representations specify how multiple overlapping object histories are understood as a whole.

A comparison of various types of transfer events can serve to illustrate which aspects of event representations are captured by the basic idea of IOH, and which are omitted in argument-time structure analyses like in (1) (see the previous chapter). In the events expressed by the sentences (a*) Laura handed the ball to her mother* and (b) *Laura threw the ball to her mother*, histories are specified for the object concepts *Laura*, *her mother*, and *the ball*, which overlap in time and space: There is an initial phase in which *Laura* has possession of the ball and a subsequent phase in which *her mother* has obtained possession of the ball; meanwhile, the ball moves from *Laura* to *her mother*. Whereas (a) and (b) are uniformly captured as a structure with three time intervals by ATS (which together constitute a two-stage event structure simply defined by the 'before' and 'after' obtaining possession), IOH accounts for every intermediate state between the initial and final phase. This fine-grained approach has relevance for predictions about the properties of the cognitive representations of (a) and (b). While the change in possession in (a) occurs at once (the moment *Laura* gives up possession, *her mother* obtains it), (b) involves a transitional phase with a trajectory of *the ball* flying through time and space before reaching a final possessor. By including this trajectory of state changes, it can be explained that (a) and (b) bring about different conditions for the success of the transfer. Specifically, this concerns the inference that change of possession is implied in (a) but need not be in (b). Semantic conflicts in linguistic descriptions of these events serve to test this intuition: Whereas (c) *Laura threw her mother the ball, but it got caught in the wind and it did not reach her* seems plausible, (d) *Laura gave her mother the ball*, but she never got it is less easy to imagine.

In IOH, it is assumed that the most basic level of perceiving events involves the activation of multiple states of one or more objects, i.e., their entire history of changes. For the perception of events, this means that not only all object concepts, but every state of these objects (i.e., not only object tokens, but object state tokens) are activated from, e.g., a linguistic or a visual input. Furthermore, multiple perceived object states are integrated into coherent object representations and mapped onto generalized, abstract schemata

retrieved from semantic memory. This means that in IOH, thematic roles do not form a basis for event comprehension but are instead inferred from perceived object changes.

3.1.4 Object representation and cascades

The presented theories ATS and cascade theory constitute different approaches to the description of event representations. If we compare the argument-time structure and the cascade of the verb *give*, we see that whereas the argument-time structure specifies an abstract schema in (1), the cascade (2) describes a tokenization of abstract act-types. The act-token underlying the A' may imply, as in this example, an action of someone transferring an object to another person per hand. However, the act of giving could in principle involve other manners of performing the action. Note that in the context of cascade theory, it is not sufficient to describe an action in time and space as 'x performs an action', since, in that case, it would not describe an act-token, but an act-type (hence, only an act-token can instantiate multiple act-types at the same time).

If we compare the inner structure of the argument-time structure in (1) and cascade in (2), we see in (2) in contrast to (1) a bifurcation of actions.

(1) Laura gave (/handed/passed/sent/mailed) a letter to her mother.
ACTIVE <X, t1> & NOT-HAVE <Z, Y, t2> & HAVE <Z, Y, t3>

(2) A' x causes y to obtain possession of z
↥
A x causes z to go to y by handing it

In (2), the bifurcation of actions is necessary in order to represent the cascade. The reason that not only one level is sufficient is that one has to perform an action to change a possession relation. This means that the action of giving (i.e., causing someone to obtain possession) implies an action of, e.g., physical movement at a lower level. This could also be just the utterance of a sentence that expresses that someone wishes or permits someone

to receive something, such as ‘x says “here you go”’. Thus, the understanding of the changed possession relation is based on an activity of the agent (x), which for instance could be causing an entity to move to a goal by handing it. At this point, the assumptions of the two theories overlap: There is a time (cf. ATS) or an action (cf. cascade theory) in which an agent is active and performs an action. In addition, there is a subsequent, separate time, at which a change of state is caused. Hence, both ATS and cascade theory separate the activity of the agent from subsequent time intervals denoting a change of possession.

3.2 Experimental evidence for temporally complex event structures

A fundamental principle of IOH is the assumption that a state change is only recognized as such if multiple object states are activated. If we consider an example *Laura opened the door*, the door can only be understood as an affected entity of an opening-action if it at the same time is understood that the door had to be closed prior to the action. In other words, for a door to be opened, it needs to be closed beforehand. The idea here is that in order to understand that the door is an affected entity changing state, the before- and after-states need to be activated at the same time; otherwise, the door would not be understood as part of an event, but merely as a state (‘being open’). The idea that multiple object states form the basis for understanding an event is captured by ATS. In ATS, however, this principle was described as counterfactual relations between states: The opening of a door is counterfactually related to an earlier, implied state of the door being closed.

In what follows, we will review research that examines, first, the hypothesis that event comprehension involves activation of multiple object states and, second, the mechanisms involved in predicting objects and their states during the unfolding of an auditory, linguistic input.

3.2.1 Competition between activated object states

In IOH and ATS, it is assumed that understanding an event is presupposed by the activation of multiple object states or even the entire history of one or more objects. This idea generates the prediction that multiple object states are activated and compete with each other when a contextually relevant object state needs to be retrieved, a prediction that was first tested by Hindy et al. (2012). By measuring brain activity (in fMRI experiments), the authors found activation patterns indicating that object states are simultaneously retrieved and compete with each other. In this study, experimental participants read pairs of sentences describing two different events involving the same affected object and had to decide whether the two sentences formed a coherent discourse or not. The first sentence described an affected object that was either minimally changed (*The squirrel will sniff the acorn*) or as substantially changed (*The squirrel will crack the acorn*). The second sentence described either a preceding or a following interaction with that object (*But first, it will lick the acorn* or *And then, it will lick the acorn*). While understanding the second sentence, it was assumed that participants prior to this (i.e., while hearing the first sentence) would have activated a state of the acorn either as intact, i.e., having undergone a minimal change (being sniffed) or a substantial change (cracked). It was expected that an acorn would be more difficult to image being licked in its substantially changed state compared with the minimally changed state. The experiment was set up to measure the semantic conflict that would be induced upon retrieval of a contextually relevant object state. The results of the study showed activation patterns indicative of stronger semantic conflict when processing sentences with substantial object changes compared to sentences with minimal object changes. Further experiments confirmed that the effects were not due to the processing of specific lexical items (Experiment 2 in Hindy et al., 2021; Solomon et al., 2015). Moreover, results from two other studies suggest that the competition effect is associated with subsequent reference to the object (Kang et al., 2020a; Prystauka, 2018).

In studies that have measured reaction times in picture-sentence matching tasks (Kang et al., 2020b; Horchak & Garrido, 2021), it has been shown that after a sentence that implies a change in an object state, initial and final object states remain activated even if

the object is not mentioned again in the second sentence (see also Hindy et al. (2015) for fMRI studies of the neural mechanisms underlying activation patterns of this type).

In experiments with reading times and reaction times in lexical decision tasks, Gennari and Poeppel (2003) found that verbs denoting multiple object states require longer processing times compared to verbs denoting a single state, which according to the authors is due to the more complex event structures of this verb type. Gerwien (2011) similarly compared reading times for intransitive verbs that do or do not denote a state change and also found that more complex verb meanings require longer processing times.

3.2.2 Predictive processing of objects and their states

Research on prediction mechanisms in language processing has examined how the comprehension of different sentence elements can drive eye-movements to visual objects. Studies using sentences with manipulated verb meanings in the visual-world paradigm (Cooper, 1974; Tanenhaus, Spivey-Knowlton, 1995) have tested whether eye-movements could be observed to referents that satisfy semantic selectional restrictions of the verb as well as conceptual world knowledge about typically associated referents of the verb. Altmann and Kamide (1999) showed that participants fixate a concurrently visible edible referent (a cake) more often than a non-edible one (e.g., a toy car) when listening to a sentence like The boy will eat the cake. The effect was already observed after the onset of the auditory presentation of the verb and before the onset of the theme (the cake).

Kamide et al. (2003) showed that while listening to a sentence such as *The man will ride the motorcycle* during free inspection of a visual scene containing a motorcycle and a carousel, i.e., multiple referents that met the verb's (*ride*) selectional restrictions, participants more often fixated on the referent (*the motorcycle*) that was assumed to be typically associated with the sentence's agent referent (*the man*). The authors argued that parsing sentences requires the integration of conceptual world knowledge in addition to semantic knowledge. Altmann and Kamide (2009) further raised the question of whether, in addition to predicting upcoming referents, listeners also predict the different states of the referents. Participants were presented with scenes containing, e.g., a woman, a wine

glass, a bottle of wine, and a table. The scene then disappeared, leaving a blank screen, and participants heard a sentence defining a context in which an object (the wine glass) either had been moved or not, such as *The woman will put the glass on the table* or *The woman is too lazy to put the glass on the table*. This was followed by a target sentence *She will then pour the wine into the glass* describing an interaction between the agent (*The woman*) and the theme (*the glass*) from the context sentence. The results show that after hearing a context sentence describing that the object had been moved to a new position (e.g., the table), and while hearing a target sentence defining an interaction with the moved object, participants were more likely to fixate the area of the blank screen where this new position of the object had been shown, compared to its original position in the scene. This thus reflects that mental representations generated through the linguistic input were mapped onto representations of previously seen objects. Importantly, fixations on the new moved position already increased significantly after verb offset (*pour*). Together, these studies show that plausible referents *and* their states (i.e., locations) are predicted during incremental sentence comprehension.

The mental capacity for prediction in perceptual processing can be attributed to a representation that is temporarily stored in working memory and continuously updated as perceptual input interacts with event schemas from long-term memory (Zacks, Speer, Swallow, Braver, & Reynolds, 2007; Zacks & Tversky, 2001; Radvansky & Zacks, 2014). By associating current perceptual experience with generalized event schemas, an intermediate mental representation (a so-called working model) allows us to understand the world as it unfolds around us and, based on past experience, to predict what will usually happen next in a given situation.

4 The interaction between language, events, and cascades

In the previous chapters, approaches capturing the multilayered and temporal complexity of actions and events were outlined and applied to cognitive representation. The analysis especially focused on the notion of cascades, which can be seen as an aspect of event representation. In this chapter, it will be examined how actions and events are verbalized. We will turn to a phenomenon which is central to linguistic encoding of event representations, namely argument structures. We will examine literature that describes the semantics of argument structures with comparison to the semantics of the verbs that frequently occur in them. Next, we will explore the nature of the cognitive representations that form the basis of linguistic encoding. This will be related to notions of multilayered and temporally complex structures as described in chapter 2 and 3.

4.1 Argument realization

A fundamental question in lexical semantics concerns the nature of the structures and contents that constitute the meaning of a word. Talmy (2000) distinguishes two subsystems in the cognitive representation of lexical concepts – a lexical and a grammatical – both of which contribute to the meaning of language. The grammatical subsystem specifies schematic content and thus has a structuring function. In a complementary way, the lexical subsystem specifies the richer content and thus fills in 'gaps' of more abstract structures. According to Evans (2010), schematic and content-rich aspects of lexical concepts are instantiated by elements of the closed and the open part-of-speech classes, respectively. In (4), the former word type class is marked in bold.

(4) **The** president **has** hand**ed the** winner **the** trophy.

The forms **has** and **-ed**, which are associated with the grammatical subsystem, indicate to the listener through the marking of the present perfect aspect that the handing-event took place before the time of utterance and is now completed. The active declarative word order indicates that the subject (*The president*), the first post-verbal object (*the winner*), and the second post-verbal object (*the trophy*) encode the agent, recipient, and the theme of the handing-event, respectively. The forms *president*, *winner*, and *hand* belong to the open word class, with the former two nouns specifying objects entities and the verb *hand* specifying an event (see chapter 2).

The basic semantics of (1) can be paraphrased as a more abstract scene: Someone causes another person (or an entity capable of possession) to obtain possession of something. We additionally see a scene with this semantics in (2), where the three referents – *the president*, *the winner*, and *the trophy* – are expressed in a parallel manner, but in a different sentence form:

(5) The president has handed the trophy to the winner.

Linguistic theories offer different approaches to explaining this variance of argument realizations as shown in (4) and (5). The argument realization in (4) will be referred to as a double-object construction (DO; with the syntactic structure: subject–indirect object–direct object) and in (5) as a prepositional-object construction (PO; subject–direct object–prepositional object). Thus, both sentence types (DO and PO) differ in their syntactic realization of the semantic recipient argument (*the winner*). Together they form the so-called ‘dative alternation’ (cf. Levin & Rappaport Hovav, 2008).

4.1.1 Lexicalist approaches

The analysis of the dative alternation raises several questions. First, as Levin and Rappaport Hovav (2008) have identified for the English dative alternation, there are two primary approaches to analyzing the meaning associated with each realization pattern (DO and PO). On the one hand, the lexicalist approach analyzes the semantics of each

argument realization uniformly. This assumption is particularly related to theories that analyze syntactic argument structures as relational syntactic or semantic specifications projected by the lexical stem of the verb. These include dependency-structural analyses based on Tesnière's (1959) valency theory, which is used to identify regularities of semantic arguments that are required by the verb lexeme. Valency theory assumes – without reference to any psychologization of linguistic expressions – that lexeme-based projection rules reflect ontological phenomena of relations between required and optional participants of actions and events.

However, forerunners of Tesnière's theory had already put forward the idea that the verb structures the sentence through its connection to other sentence elements. Comparably to the concept of valency, Karl Bühler (1934: p. 172f.) speaks in his work *Sprachtheorie* of a principle of 'elective affinities' (*Wahlverwandtschaften*) between words. According to Bühler, the verb opens up 'gaps' around itself whose occupation is limited to words of other set word classes.

A similar idea had also already been formulated by Gottlob Frege (1879) in his work *Begriffsschrift*. According to Frege's predicate logic, the predicate is the part of an expression whose truth value can be determined. The assessment of truth value can be made on the basis of one or more arguments which are analyzed under the concept denoted by the predicate. The aim of Frege's predicate logic was to grasp language in terms of mathematical values; this basic idea was adopted by valency theory within the philosophical framework of structuralism with a language-descriptive interest in mind (see, e.g., Zalta, 1995). Later, in psycholinguistic research, this basic idea was implemented by assuming that verbs provide access to a generalized "situation structure" that specifies semantic properties of the entities that fill the thematic roles of the verb (see, e.g., Ferretti, McRae & Hatharell, 2001).

Together, Fregeian and Tesnièreian theories oppose Aristotelian propositional logic, which is based on a bipartite subject-predicate structuring of propositions (see, e.g., Smith, 2020). From the bipartition follows a juxtaposition of subject and predicate, and the truth value of an expression is judged through a statement (expressed by the sentence predicate) about an object (the sentence subject), not from the predication of potentially

multiple arguments that are subordinate to the verb. The Aristotelian binarity principle manifested itself in grammar models contemporary with Tesnière's valence theory, whose analysis is based on constituency, i.e., phrase structures. The constituency-based approach analyzes structures (e.g., argument structures) in binary branching constituent relations. This principle was adopted by, e.g., transformational grammars (see Chomsky, 1993; 1986 *Generative Grammar*; Kaplan & Bresnan, 1982 *Lexical Functional Grammar*).

However, grammar theories of the lexicalist approach, which analyze either dependency or constituency, share certain basic criteria for the analysis of linguistic structures. First, they are projectionist in that syntactic structures are analyzed as projections of lexemes; projected structures are thereby exclusively components of lexical contents. Second, they follow the criterion of truth value in defining semantic content. Different patterns of realization (e.g. of the dative alternation) that share the same truth value are considered semantically equivalent. Third, they analyze linguistic expressions as compositionally transparent structures. This is closely related to the assumption of rule invariance, in which a delimited 'core grammar' is implemented to explain an ideal grammatical competence and is taken as valid for any linguistic phenomenon. This means that exceptions that cannot be captured by a core grammar need to be largely ignored (see, e.g., Stefanowitsch, 2011).

4.1.2 Constructional approaches

An alternative to the lexeme-based approach, which is here labelled as the constructional approach, assumes that each syntactic realization of, e.g., the dative alternation, has a unique meaning associated with it (Green, 1974; Pinker, 1989; Gropen et al., 1989; Goldberg, 1995; Beck & Johnson, 2004; Krifka, 2004; Rappaport Hovav & Levin, 2008; Beavers & Koontz-Garboden, 2020). If we follow this approach, questions arise about the systematics by which the meanings associated with DO and PO realizations differ. In most analyses of PO structure, a (partly) consistent semantics is identified, which in a first step can be defined following Goldberg (1995): An agent causes a theme to move

along a path to a goal (cf. Gruber, 1965; Jackendorf, 1972). Here, the path defines a more or less concretely spatial component. This also includes, for example, changes of location in a more abstract, virtual space as in *I sent an e-mail to her*. In the case of the DO structure, Goldberg (1995) defines its meaning as a more abstract possession relation that does not associate a spatial component: An agent causes a recipient to possess or simply receive a theme. Hence, this basic semantics is evident in both *give me the book* as well as *tell me a story*.

However, the determination of unique semantics associated with each of the respective realization patterns does not come without challenges. For example, the semantically synonymous (but information structurally distinct) argument realizations in *Laura gave her mother a letter* (DO) and *Laura gave a letter to her mother* (PO) pose a problem for semantically distinguishing the DO and PO structures: In the dative alternation of *give*, there is no apparent semantic difference, since both realizations seem to denote the same caused possession meaning. Problematic cases for the assumption of unique semantics for each realization pattern will be presented in chapter 4.1.3. by using examples from German and English. We will focus primarily on approaches by Goldberg (1995; 2006), Levin and Rappaport Hovav (2008), and Beavers and Koontz-Garboden (2020), which share the assumption that argument structures (e.g., the DO and PO), i.e., abstract syntactic structures, are associated with semantics that are unique to each realization pattern. Each of these approaches more or less explicitly assumes that argument structures are cognitively represented as lexically abstract units (i.e., in independence of any lexical item projecting them). However, these approaches differ, on the one hand, in the assumption of the extent to which argument realizations must necessarily be selected by the verb that instantiates the argument structure, and, on the other hand, in how sharply the semantic contributions of verb and argument structure are to be divided.

4.2 Construction Grammar

Among the constructional approaches to language (see the previous chapter), Construction Grammar has emerged as a model for grammatical analysis. It is primarily based on

identifying meanings that are associated with syntactic forms and which can be identified across the verbs that occur in them. Focusing on the English dative alternation, we will take a closer look at the assumptions that come with the Constructional Grammar approach to language in this chapter. Principles of compositionality and what meaning syntactic forms contribute relative to verb meaning will be scrutinized. We will then look at the evidence from theoretical linguistic analyses of dative alternation for and against Construction Grammar principles. This will form the basis for analyzing how the idea and cascades relate to principles assumed by Construction Grammar and lexical semantics (see chapter 4.3).

4.2.1 Preliminary assumptions

The approach to the analysis of semantics associated with a particular syntactic realization pattern can be traced back to Fillmore's (1968) Case Grammar. Fillmore was the first to introduce the role of semantics in the analysis of grammatical structures, which was later adopted by various approaches to Construction Grammar (see in particular the cognitive orientation of Construction Grammar by Goldberg, 1995; 2006). The starting point for the Construction Grammar analysis of linguistic structures relates to Fillmore's assumption that grammatical structures can be understood as so-called case frames. Case frames are schematically organized perceptual experiences with (e.g.) actions and events, that humans have repeatedly experienced in the world. As we saw in chapter 3.2, Fillmore's theory suggests that such experiences are generalized as action types that can be described by the meaning of a specific verb. For example, a cutting-frame associates the type-specific roles 'someone who cuts' and 'something that is cut'. In addition, experiences can be captured as more abstract generalizations that go beyond what can be expressed by a single verb. The action type underlying the case frame of a cutting-action could be captured as a transitive (i.e., causative) schema in which someone, at a more abstract level, brings about a change of state from an affected entity by performing an action. For the sake of simplicity, such frames are referred to as constellations of thematic roles such as agent, theme, and recipient.

Scenes, which are stored as mental schemata and can be retrieved during speech production or comprehension, are thus based on generalized experiences of events involving, e.g., an agent performing an action, and a patient or theme being acted upon. Central to Fillmore's assumption of abstract, generalized scenes is that experiences are not merely assembled as meaning components (which we categorize here as thematic roles), but instead, they are perceived and uniformly represented as wholes (i.e., non-compositional structures).

If we turn our attention to the role of grammar, Fillmore's theory predicts that specific syntactic structures associate these scenes. From now on, these associations between syntactic structures and scenes will be called constructions. The DO construction (or the syntactic argument structure 'subject–indirect object–direct object') associates a conceptual frame, relative to which conceptual roles of involved referents are defined. Goldberg (1995: p. 141) refers to the meaning of the scene associated with the DO construction with the terms *caused possession* or *transfer of possession*. An event of this type specifies an entity, the agent, who performs an activity with the intention of causing another entity, the recipient, to obtain possession of a third entity, the theme – in short, X has the intention to cause Y to possess Z. The role of each entity involved in the transfer event is defined relative to the overall transfer scene. Langacker (1987: p. 32) describes such scenes as cognitive domains. To illustrate the idea, he uses the example that a hypotenuse can only be defined relative to a right triangle, and an elbow can only be defined relative to an arm. This analogy serves to suggest that meaning components have to be relativized to an overall scene, that is, interpreted contextually.

4.2.2 The constructionist approach to the dative alternation

Theories such as those of Fillmore (1968) and Langacker (1987) (see also Gruber, 1965; for theoretical foundations of the cognitive representation formats schemata, scripts, and frames, see Bartlett, 1932; Schank & Abelson, 1977; Minsky, 1975) assume that meaning is relativized to holistic representations of generalized experiences. It follows from these approaches that grammatical structures delineate such representations from each other. In

subsequent research aimed at the role of semantics in grammatical analysis, Lakoff's (1977) and Goldberg's (1995; 2006) Construction Grammar in particular has formed a framework for explaining meanings associated with syntactic argument structures. According to the Lakoff-Goldberg approach, generalized scenes, i.e., semantic structures, are mapped onto syntactic structures. In this, linguistic structures are conceived as cognitive units of syntax-semantics associations. According to the Construction Grammar approach, these units thus form constructions of form and meaning. In the following, the notion of a construction will be applied in a two-fold way. On the one hand, the notion refers to syntagmatic relations in argument structures (on a syntactic or semantic level). On the other hand, it will, in line with the sense used in Construction Grammar, refer to units which 'construct' a relation between form and meaning. A unification of form and meaning implies the prediction that a change in syntactic structure goes hand in hand with a change in semantic structure.

The Construction Grammar approach differs in many ways from lexicalist, projectionist grammars (e.g., Chomsky, 1993; 1986 *Generative Grammar*; Kaplan & Bresnan, 1982 *Lexical Functional Grammar*). Construction Grammar arose from the desire for a comprehensive or complete coverage of linguistic phenomena within a single theoretical framework. To achieve this, a theoretical framework must be able to analyze linguistic phenomena both within and outside the core grammar. Thus, both compositionally transparent and non-compositional, intransparent structures are to be explained by the same principles. Instead of assuming the definition of a small set of invariant rules derived from an idealized grammatical competence, representatives of the Construction Grammar approach take idiosyncrasies that arise from linguistic performance into account. In particular, linguistic phenomena of the semantically intransparent type (hence, deviating from the core grammar) are given a central role in order to demonstrate advantages of the Construction Grammar analysis.

If we turn to the argument realization patterns of the dative alternation, semantically transparent and intransparent expressions can be observed. The first category includes

prototypical[4] instantiations of the dative alternation as in (6) and (7). In the following chapter, we will examine Construction Grammar principles of compositionality on the basis of intransparent expressions.

(6) The President handed the winner the trophy. (DO)
(7) The President handed the trophy to the winner. (PO)

Following a decompositional analysis by Goldberg (1995) and Jackendoff (1983) (see also, e.g., Harley, 2003; Krifka, 1999, Pinker, 1989), syntactically relevant aspects of the meanings of (6) and (7) can be captured by the following simple paraphrases:

[4] Prototypicality is understood here in two respects: First, the argument structure projected by the verb lexeme instantiates a sentence structure that can be exhaustively analyzed by verb semantics and is thereby compositionally transparent; the argument structure of the verb is in this sense a prototype of the sentence. Second, following Lakoff (1977) and Goldberg (1995), prototypical meaning is understood as a 'central' meaning of the verb. The general applicability of prototypical verb meanings is not discussed here; instead, they are adopted for the sake of simplicity.

(8) X has the intention to cause Y to receive Z (DO)

(9) X causes Z to move to Y (PO)

Thus, both from a Construction Grammar and constructional approach to the dative alternation in general, it is assumed that DO and PO structures must each be associated with different semantics and form independent constructions of form and meaning. According to Goldberg (following Lakoff and Fillmore), the simple paraphrases (8) and (9) capture a generalized scene, which are associated with the syntactic structures.

The one hand, the lexicalist approach assumes that both syntactic realizations of the dative alternation share a single, common meaning. This single meaning representation comprises a thematic structure with an agent, subject, and recipient which is consistent across syntactic structures, i.e., has a shared truth value across realization patterns. Moreover, this is based on the fundamental assumption that the same event can be described in multiple ways, and that syntactic differences in argument realization can be attributed to independent syntactic rules (cf. Beck & Johnson, 2004; Pesetsky, 1995). On the other hand, Construction Grammar uses a different paradigm for determining differences in meaning. In this paradigm, it is not truth values of semantic structures that are investigated, but instead differences in cognitive representations of associated event structures. Differences in meaning are thus determined 'below' the threshold of truth value, as a syntactic structure expresses a specific cognitive construal of an event:

> "Differences in semantics are not necessarily truth-functional differences, but may represent a different construal of the situation being described; that is, the relevant semantics is speaker-based." (Goldberg, 1995: p. 8).

The starting point for determining semantic representations, according to Goldberg, is the event conceptualization that occurs during speech production. This cognitive definition is also reflected in Pinker's (1989) differentiated meaning representation of DO and PO structures, according to which the dative alternation expresses a "conceptual gestalt shift".

Together, constructional analyses reject explaining argument realization variations as syntactic operations that are processed independently of associated semantic or cognitive representations (cf. Chomsky, 1993; 1986; Kaplan & Bresnan, 1982). Instead, two different accounts are used to explain the meaning representations that underlie conceptualization during language production. On the one hand, the thematic role approach (Fillmore, 1968; see also Gruber, 1965; Jackendoff 1983; 1972; Cai, Pickering, & Branigan, 2012) has been established as a perspective on semantic representation. According to this approach, the event participants expressed in (6) and (7) correspond to a specific configuration of the atomic primitives agent, subject, and recipient, which are associated with functions at the syntactic surface: Agent to subject, theme to direct object, and recipient to indirect object (in the DO) or prepositional object (in the PO). On the other hand, the Lakoff-Goldbergian Construction Grammar takes an alternative explanatory approach, assuming that event structures, rather than thematic roles, form the conceptual basis for language production (see also Jackendoff, 1998; 2002; Pinker, 1989; Rappaport Hovav & Levin, 1998). According to this approach, verb meanings (i.e., constructional meanings) are decomposed into primitive predicates (ACT, BECOME, CAUSE, HAVE) and hierarchically embedded in each other. Thus, the DO structure consists of a HAVE predicate embedded in a CAUSE predicate, while the PO structure contains an embedded BE AT predicate (cf. Levin & Rappaport Hovav, 2005):

(10) X CAUSE [Y HAVE Z] (DO)
(11) X CAUSE [Z BE AT Y] (PO)

The cognitive reality of an event-structural meaning representation of ditransitive verbs (more specifically the verb *give*) in DO and PO structures (as shown in (4) and (5)) was experimentally investigated by Ziegler, Snedeker, and Wittenberg (2018). The authors tested whether semantic structural priming is sensitive to thematic roles or event structures a sociated with different syntactic realizations. They argued that if thematic roles are primed, then different syntactic realizations of the patient of a light verb (*her son* in *The mother is giving her son a hug/...a hug to her son*) should prime the choice of syntax

in descriptions of transitive events (e.g., a picture of a mother giving her son a kiss on the cheek) more than the recipient of a ditransitive verb (*her son* in *The mother is giving her son an apple/...an apple to her son*). Since the study showed no difference in the magnitude of priming effects between light and ditransitive verbs, the authors concluded that priming was not sensitive to different syntactic realizations of the patient of light verbs, but only to syntax and its associated event structure, which for light verbs was assumed to be 'X acts on Y' across syntactic realizations. According to the authors, this means that event structures, not arrays of thematic roles, are mapped onto syntactic forms.

In summary, Construction Grammar assumes unique semantic interpretations of DO and PO structures, which on the one hand are linked to generalized perceptual experiences (or "scenes"), and on the other hand are mentally represented not only as configurations of thematic roles, but as complex event structures with embedded structures, each of which can be captured by simple predicates.

4.2.3 Compositionality by instantiation

Theories of lexical semantics (e.g., Asher, 2011; Bierwisch, 1982; Pustejovsky, 1995) strive, on the one hand, to describe lexical concepts with an invariant system of rules and, on the other hand, to account for the need to accommodate contextual interpretations of lexical concepts. Their goal is to preserve the so-called Fregeian principle of compositionality, according to which the overall meaning of an expression can be derived from the meanings of its component parts (as first formulated in Frege's *Begriffsschrift* (1879); see, e.g., Cook, 2023). In syntactic theories following the lexicalist approach, the syntactic properties of a word are considered to be independent of the meaning of the word. Syntactic rules thus determine combinations of phrases and sentences without adding conceptual content to the meaning contributed by lexical concepts nor changing combinatorial properties of words (cf. Kay & Michaelis, 2011). Jackendoff (1997: 48) describes this approach to compositionality as a "doctrine of syntactically transparent composition" that identifies the locus of the overall sentence meaning in the lexical-conceptual structures of the words that constitute the sentence.

By contrast, Construction Grammar follows the hypothesis that basic units of the linguistic system consist of symbolic associations of phonological and semantic and pragmatic information (Goldberg, 1995; 2006; Langacker, 1987; 2009; see also Evans & Green, 2006). The basic assumption of Construction Grammar approaches is that these basic units – constructions – are conventional pairs of forms and meanings at different levels of abstraction and complexity. At the core of the Construction Grammar approach is Fillmore's (1968) assumption that meaning components are to be understood relative to the overall meaning of an expression (see also Goldberg's (1995) *scene encoding hypothesis*). Hence, sentences are analyzed not only by their constituent words and their syntactic and semantic properties, but they are form-meaning constituting units in themselves.

For Goldberg (1995; 2006), the concept of constructions refers to abstract syntactic argument structures or argument structure constructions. While, e.g., Dependency Grammar (cf. Tesnière, 1959), Generative Grammar (e.g. Chomsky, 1993) and Lexical Function Grammar (Kaplan & Bresnan, 1982) divide form and meaning into atomic and modular linguistic components, Construction Grammar does not regard linguistic form independently from meaning, but instead examines variations in meaning associated with formal variations at the syntactic surface. Thus, the interest of Construction Grammar is not an onomasiological analysis of cognitive representations and their potential forms of expression, but a semasiological analysis of linguistic forms and the associated, cognitive representations that are specific to these forms.

An argument for this semasiological approach relates to a hypothesis of language acquisition from Tomasello's (1992; 2002) *verb island hypothesis*, which predicts that verb-specific argument structures learned during first-language acquisition are gradually generalized across verbs as abstract syntactic constructions. For example, it is assumed that similarities between specific participant roles of verbs such as *kick* (associating a kicking and kicked entity) or *throw* (associating a throwing and thrown entity) are recognized and stored as the generalized participant roles agent and theme, which are shared among verbs (see also Diessel, 2013; Ellis, 2013 on Construction Grammar research in first- and

second-language acquisition; for language production, see Bencini, 2013 and for neurolinguistics, see Pulvermüller, 2013).

To identify a construction, Goldberg (1995: p. 4) introduces a definition that focuses on the compositionality of the construction. In this definition, a linguistic pattern (such as the DO or the PO realizations) can be considered a construction of form and meaning if no aspects of its form or meaning are predictable from properties of its constituents. This mainly concerns that if argument structures are not predictable from the valency of the verbs occurring in them, then their composition cannot be transparently analyzed by looking at the verb on its own. In Goldberg's later work, the principle of compositionality is modified and extended to the cognitive dimension, with fully compositional expressions attaining the status of a construction when they consolidate as units through frequent use (i.e., when they are entrenched as chunks) (Goldberg, 2006: p. 1–4). In the example of the PO construction, the relevance of this definition of compositionality is made evident by the implementation of idiosyncratic verbs, as in the classic example (13) from Goldberg (1995: p. 29):

(12) Sam pushed the napkin off the table.
(13) Sam sneezed the napkin off the table.

The event structure X CAUSE Z BE AT Y can be interpreted from (12): X (*Sam*) causes Z (*the napkin*) to change spatial position (from on the table to not on the table) by performing the act of pushing. In (13), a similar event takes place by performing the act of sneezing. Importantly, whereas the syntax can be predicted by the caused motion verb *push*, it cannot be predicted by the intransitive verb *sneeze*. This leads Goldberg to theorize that the argument structure is not projected by the verb (cf. the lexicalist approach) but instead must be mentally stored as an independent linguistic unit. The verb lexeme and the NPs of the sentence thereby instantiate an abstract argument structure construction, which in turn determines the meaning of the verb lexeme and the implemented NPs according to Fillmorean principles. For this example, Goldberg's Construction Grammar analysis focuses on the valency extension of the verb *sneeze*, which, according to her analysis, is

undergoing a coercion of meaning according to the semantic constraints introduced by the argument structure construction (i.e., the PO construction). In other words, verbs get fundamental properties of their semantics defined by the construction they occur in. According to Goldberg, in order to account for the use of *sneeze* in the PO (as in example (13)), a lexicalist approach would need to (implausibly) assume that, in addition to the prototypical intransitive use, *sneeze* must also be stored in the mental lexicon as a caused motion verb.

However, as both Gawron (1985) and Pustejovsky (1991) note, the CAUSE-BE AT interpretation in (13) could be explained by a pragmatic inference. Specifically, this inference involves that from the individual predicates X CAUSE (Sam sneezes) and Z BE AT Y (the napkin is no longer on the table), a logical bridge between cause and effect can be derived that does not need to be interpreted by the semantics of the expression per se. Goldbergs (1995: p. 155) argues against this objection on the one hand following Talmy (1985) by noting that the expression (13) is not permitted in every language (this would follow from the assumption that predicates could simply be combined by logical inference). On the other hand, she notes that logical inferences come about in language comprehension, while example (13) shows that a novel use of the PO construction that is available in language production (see also Kay & Michaelis, 2011).

According to Goldberg, similar valency extensions also appear from novel use of the DO construction:

(14) Sam baked her sister a cake.

In (14), the event structure X CAUSE Y HAVE Z is evident: By baking, Sam intends her sister to receive a cake. As in example (13), the abstract argument structure (DO) is assumed to associate an event structure that coerces a transfer interpretation of the transitive verb *bake*. One might object that this transfer meaning does not imply that the recipient necessarily comes into possession (as shown by the contextual cancelability of a successful transfer in *She baked her sister a cake, but her sister never discovered it in the fridge*).

However, according to Goldberg, the action is performed with the intention that a recipient will receive a theme.

Since the semantics of the above-described argument structure constructions can be identified across verbs, they must be regarded as type constructions that can be instantiated by verb lexemes, i.e., tokenized by specific lexical items. For Goldberg, the compositionality principle refers to the composition of type constructions, not token constructions. However, as Welke (2019: p. 29) notes, linguistic expressions cannot be analyzed with respects to their composition until they have become tokenized, because only a tokenized argument structure construction contains lexical content. And since, in turn, e.g., verb lexemes in token constructions have been made compositional by coercive mechanisms, idiosyncratic expressions can in principle be regarded as semantically transparent compositions. This is also implicitly assumed by Goldberg, since she assumes that verb lexemes receive a new meaning by being implemented in the construction.

To accommodate this observation, Welke replaces Goldberg's principle of compositionality with the assumption that constructions acquire their independent status by a conventionalization criterion alone. This means that the mental *construct-i-con* – the representational network of each construction of a language in long-term memory (from morpheme to sentence level) – exclusively contains conventionalized pairs of form and meaning. Hence, this differs from Goldberg's criterion that a construction is identifiable by its non-compositional structure. The conventionalization criterion covers, on the one hand, schematic type constructions (e.g., argument structure constructions) and partially schematic (or partially lexically filled) constructions that have not been tokenized and therefore cannot be assessed with respect to their internal compositionality, and, on the other hand, token constructions or lexically fully specified and idiomatic constructions that are stored in long-term memory as non-compositional structures.

It is argued here that Welke adopts a processual perspective for his definition of compositionality: During the mental retrieval and composition of argument structure constructions and the lexical concepts instantiating them, compositional representations emerge. The verb lexemes in (13) and (14) thus acquire a compositional meaning by

coercive mechanisms. Welke's definition is in principle consistent with Jackendoff's (1997; p. 49) notion of a construction-based compositionality:

> "To embrace a construction-based model of semantic composition is not to reject the existence of syntactically transparent composition but instead to treat it as a 'default in a wider array of options'.".

Michaelis (2017; p. 12) makes a similar argument:

> "To take a construction-based perspective on semantic composition is not to deny the existence of syntactically transparent composition: if a class of expressions can be represented by means of a phrase structure rule that is paired with a rule that composes the semantics of the mother from the semantics of the daughters, 'a construction-based approach would propose a construction that is functionally equivalent to such a rule-to-rule pair' (Kay & Michaelis (2012). What distinguishes constructional approaches is the ability to represent linguistic structures […] in which the meaning of a phrase cannot be attributed solely to the meanings of its daughters. Approaches admitting only syntactically transparent (or 'rule-to-rule') composition lack this ability."

Among the approaches that analyzes linguistic expression exclusively as syntactically transparent compositions is Pustejovsky's (1995) *Generative Lexicon* (see also Pustejovsky, 1998; 2012 *co-composition*). Contrary to the Construction Grammar approach, Pustejovsky analyzes contextual coercion of lexical concepts in a way that preserves their compositionality. In order to preserve compositionality in expressions that deviate from an assumed central sense of a word, i.e., that show a flexible meaning, Pustejovsky introduces the notion of a *qualia structure* of a lexical concept. As a part of a qualia structure, a *quale* denotes, among other things, a prototypical functionality or purpose of an entity denoted by a word. This particular quale is defined as a "hidden event" which is a part of the lexical representation (see Pustejovsky & Anick, 1988). For example, a quale of the word *bread* is a hidden baking-event. This explains why *fresh bread* can be understood synonymously with *freshly baked bread*, since the modification interacts with the baking-quale of *bread* (cf. Pustejovsky & Jezek, 2016: p. 3–5). The qualia structure of the lexical concept *bread* thus includes a meaning component that captures how the entity comes about in the world. This would also explain why the phrase *a*

baked bread seems odd, since this modification does not contribute any information not already provided by the hidden event.

However, as Bücking and Maienborn (2019) note, qualia structures do not always accurately describe contents of lexical concepts: while the modification *fast* in *a fast highway* and *a fast car* reveals a telic quale (i.e., a quale that specifies the function of an entity) of highway and car, since, e.g., a car is designed to drive and therefore can drive fast (cf. Pustejovsky, 1995: p. 413), this is not true for *a fast dog*. This is because there is no telic quale associated with *dog* that specifies that a dog's existential purpose is defined by its movement or its ability to run. Therefore, in order to grasp the qualia structure of *dog*, further experience-based world knowledge would have to be integrated, potentially leading to an inflation of world knowledge. Inflation is a problem for a robust theory of lexical representation, since one would have to continuously include new concepts to explain contextual occurrences of a word, which can lead to an implausibly extensive amount of concepts needed to explain a word's central meaning (see also Herweg, 1989: p. 106). To avoid an inflation of world-knowledge in lexical representation, Bücking and Maienborn instead propose to explain contextual interpretations by a pragmatic enrichment component (cf. Dölling, 2003). Hence, according to their solution to the problematic case of *a fast dog*, they assume that the interpretation of nominal modification is influenced by some kind of contextual element which is not part of the lexical representation.

Bücking and Maienborn thus approach the problem of adaptive interpretation of modified nominal concepts from the listener's perspective. However, following Goldberg (1995), it can be objected that the examples analyzed by Bücking and Maienborn are not solely a matter of interpretation, but also a phenomenon of generative language use from the speaker-based perspective. That is, a language-receptive approach to contextual readings does not necessarily explain what mechanisms of compositionality are available to the speaker; arguably, it is first and foremost the speaker-based perspective that deals with syntactic and semantic combinatory properties of nouns and their modifiers or verbs and their arguments.

The speaker-based perspective on the phenomena of coercive mechanisms is present in Construction Grammar approaches such as those of Goldberg and Michaelis. In

Goldberg's approach, coercion comes about through instantiation of token constructions. Schema (15) illustrates the interaction between verb and construction in the instantiation of the verb hand in the DO construction (taken from Goldberg, 1995: p. 51).

(15)

Sem	CAUSE-RECEIVE	<	**agt**	**rec**	**pat**	>
	|R		|	|	|	
R: in-	HAND	<	hander	handee	handed	>
stance,	↓		↓	↓	↓	
means						
Syn	V		SUBJ	OBJ_1	OBJ_2	

The top line in (15) defines the semantic component of the DO construction as a cause-receive meaning, which is an abstract thematic structure with an agent (agt), recipient (rec), and a patient (pat).[5] [6]

The second line contains the specific participant roles of the verb, which are fused with the generalized constructional arguments or argument roles. According to Goldberg, this fusion follows certain principles. On the one hand, a coherence criterion (*The Semantic Coherence Principle*, Goldberg, 1995: p. 50) presupposes that the verbal participant roles can be imagined as instantiations of the more abstract argument roles of the construction (hander as agent, handee as recipient, handed as theme). On the other hand, a

[5] 'Patient' will be replaced by 'theme' in the present analysis because, unlike a patient, this entity is not characterized by an intrinsic change in state, but only by an extrinsic change in spatial position (cf. Dowty, 1991).

[6] According to Goldberg, the DO construction associates cause-receive rather than, e.g., cause-possession because this meaning captures a broader variation in meaning (i.e., a greater variation in the implementation of lexical concepts): *Mary gave John a hug/kiss/an idea* does not denote a change in possession but only that someone receives something (see Beavers & Koontz-Garboden, 2020: Chapter 3 for an in-depth analysis of ditransitive verbs).

correspondence criterion (*The Correspondence Principle*, ibid.) requires that profiled (i.e., valency-bound or obligatory to the associated scene) argument roles and participant roles be fused (cf. Fillmore, 1977; Langacker, 1987). In case of discrepancy between the number of obligatory argument roles and participant roles, the construction decides (i.e., coerces) the reading of the verb:

> "Each participant role that is lexically profiled and expressed must be fused with a profiled argument role of the construction. If a verb has three profiled participant roles, then one of them may be fused with a non-profiled argument role of a construction." (ebd.: p. 50).

That is, profiled argument roles of the construction must be realized; participant roles of the verb, on the other hand, must not. This is visible in (16) by the valency expansion of *bake*:

(16)

Sem	CAUSE-RECEIVE	<	**agt**	**rec**	**pat**	>
	|R		|	¦	|	
R: in-	BAKE	<	**baker**	bakee	**baked**	>
stance,	↓		↓	↓	↓	
means						
Syn	V		SUBJ	OBJ_1	OBJ_2	

The dashed line in (16) indicates that the argument role recipient (rec), marked in bold and thus profiled, induces the conceptualization of a non-profiled (i.e., not associated with the lexical concept) participant role bakee. According to Goldberg, compositionality by coercion thus occurs top-down from the level of construction to the level of the lexical

verb (but see Michaelis, 2017). Following Welke, it can be argued that surface realizations have been made compositional.[7]

In summary, the central contribution of Construction Grammar approaches to the semantics of argument structure is the theory that patterns of argument structure (argument structure constructions) exist independently of lexical argument-projecting predicates. Goldberg's constructional analysis hypothesizes that variants of abstract constructional meanings can be instantiated by a fixed verb meaning. It thus assumes that the finite verb merges with a construction whose semantic contribution overrides the semantics of the verb via the principle of coercion in case there are inconsistencies between verb and construction. As evidence for this, Goldberg's analysis includes the PO and DO constructions

[7] It is important to note that word order variations are not specified in this model; they come about through the 'inheritance' of more abstract constructions that specify propositional or interrogative sentences or information structural variations (see, e.g., Lambrecht, 1994). This point is beyond the scope of the present description of the basic Construction Grammar principles of compositionality. Readers are referred to chapter 2 in Goldberg (1995) and chapter 5 in Welke (2020) for an alternative hypothesis in the context of German.

that cause valency-augmenting coercion when intransitive or transitive verbs that do not in themselves denote CAUSE-HAVE or CAUSE-BE-AT predicates are implemented. Compositionality is thus achieved through valency augmentation of verbs that appear in argument structure constructions that are not projected by the verbs. The basic assumption of Lakoff-Goldbergian Construction Grammar is that a syntactic structure is associated with a generalized scene (cf. Fillmore, 1968; 1977), which entails that semantic structures are mapped onto grammatical structures that together are stored as units. This further entails that variations in grammatical structure also imply variations at the level of meaning. In this analysis, the dative alternation does not only involve variations in the linearization of the same semantic structure, but the association of distinct event structures containing embedded predicates.

4.2.4 The interaction between verb and construction

As we saw in the previous chapter, Goldberg (1995) assumes that there are fusions of argument roles of the construction and participant roles of the verb, whose semantic representations can overlap either partially or completely. The explanatory potential of Construction Grammar lies especially in the instantiations of verb and construction meanings that overlap only partially. It is assumed for these cases that although the verb instantiates the construction together with other lexical concepts, the construction coerces a particular reading of the verb. This implies that the construction must have an independent mental representation as a valency-specifier that top-down projects its arguments onto the

argument structure of the verb.[8] Idiosyncratic uses of verbs (valency-augmenting instantiations in argument structure constructions) are evidence of the independent existence of abstract argument structure constructions, since otherwise multiple implausible verb meanings would have to be assumed.

This approach differs from lexicalist approaches (cf. chapter 4.1.1) such as valency theory (Tesnière, 1959) or the X-bar-theory of Generative Grammar (Chomsky, 1993), both of which assume that the governing head of the sentence projects its dependencies. In contrast, Construction Grammar, for example, opposes valency grammar by assuming that the construction ultimately decides the interpretation of the verb and not the other way around (a principle going back to Fillmore's (1968) Case Grammar). However, as

[8] One might object that the Lakoff-Goldbergian assumptions of completely overlapping instantiations is a cognitively uneconomical premise for an unnecessary double representation of both the participant roles of the verb and the argument roles of the construction (see chapter 4.2.3; see also Ágel, 2015: p. 82). However, core principles underlying Construction Grammar indeed relate to the cognitive economy of linguistic expressions: By mentally representing a limited set of argument structure constructions alongside the single entry of a verb with its prototypical valency pattern, one avoids assuming multiple entries of the same verb in order to explain idiosyncratic uses of the verb.

Welke (2009; 2019: chapter 5) and Stefanowitsch (2011: pp. 369-384) note, valency theory and Construction Grammar do not necessarily form a contradiction to each other; rather, the assumption of an interaction between construction and verb valence implies an integration of valency grammatical principles into Construction Grammar. In contrast to Lakoff-Goldbergian Construction Grammar, Welke's approaches the Construction Grammar framework by assuming an alternation between projection from the side of the construction and the side of the verb.

However, the Construction Grammar approach is challenged by other constructional approaches which are intuitively analogous to Construction Grammar (i.e., that to some extend recognize the notion of a construction), but which ultimately cannot be reconciled with Construction Grammar at fundamental points. These competing approaches, mainly represented by Rappaport Hovav and Levin (2008), identify basic meanings associated with syntactic structures. However, at the same time, they argue that the construction does not 'dominate' the verb; rather, the verb selects the construction (contra Construction Grammar).

The so-called uniform multiple meaning approach identified by Rappaport Hovav and Levin (2008) assumes that Construction Grammar must predict that every DO realization is associated with a possessive relation (Y HAVE Z) and every PO realization with a spatial goal (Z BE AT Y) (see (17)). This approach contrasts with the "verb-sensitive approach" represented by the authors, which is characterized by the associations shown in (18), and which predicts that the meaning of DO and PO realizations is always decided by the verb.

(17) Predictions of Construction Grammar (the "uniform multiple meaning approach"):

Ditransitive verbs →	caused motion	caused possession
	↓	↓
	PO	DO

(18) Predictions of the verb-sensitive approach:

give-type verbs →	caused possession	caused possession
	↓	↓
	PO	DO
send-type verbs[9]→	caused possession	caused possession
	or	
	caused motion	
	↓	↓
	PO	DO

[9] Rappaport Hovav and Levin (2008) primarily refer to throw-types (e.g., *kick*, *lob*, *shoot*, *throw*, *toss*, ...). However, they consider send-types to be equivalent with respect to CM and CP analysis. Therefore, for the simplicity of the present analysis, only the latter type will be considered.

Despite the label “verb-sensitive”, this approach (see (18)) is subsumed here alongside the uniform multiple meaning approach (see (18)) under a constructional approach (cf. chapter 4.1.2), since they can both be linked by the assumption that an argument structure can associate a specific event structure independently of verb semantics.

The verb-sensitive approach represented by Rappaport Hovav and Levin distinguishes between different aspects of the representation of a verb: first, a lexical core meaning, the *root* of a verb (Pesetsky, 1995: 70), second, the event structure that may be associated with the root, and third, the syntactic realization of the event structures. Thus, the verb-sensitive approach takes into account associations between event structures and syntax, and event structures and verbs: verb meaning can be decomposed in two parts, where on the one hand event structures are a limited set of types, and on the other hand, lexicalized roots are a potentially open number. Here, the lexicalized root is the component of verb meaning that is identifiable across all usages. Basically, give-types (e.g., *give*, *hand*, *lend*, *loan*, *rent*, *sell*, ...) are assumed to lexicalize CP exclusively (i.e., they select a recipient), whereas send-types (*send*, *mail*, *ship*, ...) do not lexicalize CP, but only CM (i.e., they select and spatial target). Furthermore, send-types can denote both CM and CP depending on (a) the argument realization (PO or DO) and (b) semantic properties of implemented arguments (i.e., whether a spatial target is an entity capable of possession). In the following chapters, we will look more closely at this idea.

4.2.4.1 Caused possession-roots as evidence for verb-sensitivity

According to Rappaport Hovav and Levin (2008), the alternations of send-types and give-types pose a challenge to the predictions of the semasiological approach in Construction Grammar. Let us first consider the alternation of the verb *give* in (13).

(19) a. She gave him the trophy. (DO)
b. She gave the trophy to him. (PO)

According to Rappaport Hovav and Levin, (19) shows that although the root of *give* only lexicalizes CP (which is consistent with the CP meaning of the DO structure and leads to a pure CP meaning in (19a)), the implementation of *give* in the PO structure leads to a synonymous, pure CP meaning (hence, without a spatial target/CM meaning). Here, it is important to note that although (19a) and (19b) could both potentially denote a spatial target, this meaning component would be contextually derivable and not lexicalized by the root (the trophy could either be handed over or simply left in a cupboard, with the change of possession recognized solely by a verbal agreement between two possessors). The lack of CM meaning in the PO realization is thus incompatible with what Construction Grammar would predict, according to Rappaport Hovav and Levin. However, as Rappaport Hovav and Levin also note (2008: p. 132 (footnote [4])), Lakoff-Goldbergian Construction Grammar does indeed assume the same semantic CP meaning for the verb *give* in both the DO and the PO realization, in line with the verb-sensitive approach.

Nevertheless, against Rappaport Hovav and Levin assumptions, it can be argued that Lakoff-Goldbergian Construction Grammar does differentiate the meaning of DO and PO realizations of *give*. The basis for this differentiation lies the idea that meaning is constituted according to a "construal principle", in which the meaning of an expression is determined 'below' the threshold of truth value: it concerns speaker-based conceptualization which is underlying the DO and PO realization of *give*. Thus, according to Goldberg (1995: p. 91), the meaning side associated with the respective argument realization does not differ semantically, but by information structural perspectivization according to the principle that last sentence argument encodes a focus component (cf. Lambrecht, 1994). Although Goldberg does not explicitly assume a spatial meaning component (CAUSE-BE AT) for the PO realization of the verb *give*, it is nevertheless reasonable to assume that Goldberg's analysis of *give* in the PO and DO realizations is based on distinguishing two different event conceptualizations that share the same semantic truth value. Hence, Rappaport Hovav and Levin's objection that Goldberg must assume a spatial meaning component for the PO realization in order to maintain the consistency of the Construction Grammar framework is inconsistent with Goldberg's own claims, since the criterion assumed by Goldberg for determining constructional meaning is located at a cognitive-

conceptual level and thus differs from Rappaport Hovav and Levin's purely semantic criterion.

However, the basic challenge for a Construction Grammar analysis of dative alternation persists: Goldberg leaves unanswered the issue that, on the one hand, the PO realization should be associated with an abstract, generalized CM meaning and thus with a spatial goal (1995: pp. 153f.), and that, on the other hand, the PO realization of *give* does not select a spatial goal. In order to preserve the semasiological principle of Construction Grammar, but at the same time to accommodate that *give* causes a CP reading of the PO realization, Goldberg differentiates the meanings of the DO and PO realizations of *give* according to the construal principle on an information structural level alone (Goldberg calls this special case "transfer caused motion": p. 91). However, this is issue is not discussed further by Goldberg.

Rappaport Hovav and Levin (2008: 138) provides evidence for the assumption that give-types do not lexicalize a spatial component, but only CP (see also Goldberg (1997) and Pinker (1989)). This comes from the fact that the PO realization of *give* cannot be interrogated in terms of any spatial properties:

(20) Where did you throw/*give the ball?

The CP meaning contains a minimal internal structure in which either the initial or final possessor is in possession, which is why modification of a more complex path is not possible:

> "Beavers (2006), Jackendoff (1983: 192), Krifka (2004: 11), and Rappaport Hovav (in press) point out that paths in transfer of possession events are two-point paths consisting of the original possessor and the recipient; they lack any internal structure. Thus, give-type verbs cannot take to phrases with modifiers further specifying the extent of the path (e.g., halfway), in contrast to throw- and send-type verbs.
> (a) *Susan gave the ball all the way/halfway to Bill.
> (b) Jake threw/kicked the ball all the way/halfway to Bill.
> (c) I sent/shipped the package halfway/all the way around the world to the Antarctic."

However, the give-types identified by Rappaport Hovav and Levin include verbs such as *hand* and *pass*, which, according to the authors (2008: pp. 136-137, footnote [8]), do indeed lexicalize a spatial component:

> "Even among the give-type verbs there is hand, which may simultaneously take a recipient and a directional phrase, as in I handed Tracy the basket over the fence. This verb lexicalizes a change of possession and in addition a change of location. The root must contain a change of location meaning component since it specifies that the change of possession is effected by hand, requiring it to be spatially realized."

In the quote above, it is evident that Rappaport Hovav and Levin recognize that the give-types *hand* and *pass* do indeed lexicalize a spatial goal. Thus, regardless of whether *hand* and *pass* lexicalize a possessor in addition to a spatial goal, the generalizability of Rappaport Hovav and Levin's assumed uniform classification of give-type verbs does not hold for *hand* and *pass*. In fact, these give-types are compatible with path modification, against Rappaport Hovav and Levin's own predictions:

(21) She handed/passed/*gave it halfway to her.

Thus, issues in explaining the dative alternation with the give-root can be identified for not only Construction Grammar, but also for Rappaport Hovav and Levin's analysis. Rappaport Hovav and Levin, unlike Goldberg, take truth value as the starting point for differentiating meanings associated with the DO and PO structure; this leads to a pure synonymy of expressions, which is why the semasiological approach of unique constructional meanings must be rejected. Rappaport Hovav and Levin's analysis thus leads to a verb-sensitive approach, where the PO realization can be associated with a spatial target (i.e., CM), but this meaning can be overwritten by a CP root. However, and problematically for Rappaport Hovav and Levin's analysis, this is not true for CP-roots such as *hand* and *pass*. Goldberg, on the other hand, assumes that although the expressions are semantically synonymous, DO and PO realization patterns express subtle differences in event conceptualizations. However, this still leaves a gap in Goldberg's explanation of the

extent to which the PO realization should in principle always be associated with a spatial target.

4.2.4.2 Caused motion-roots as evidence for the DO construction

In the previous chapter, we saw that at least the CP root give overrides any potential semantic meaning of the PO realization. On the other hand, if we consider the semantic constraints on the DO realization of CM verbs (e.g. *send*), a different picture of the semantic contribution of the construction emerges. In their analysis of send-roots, Rappaport Hovav and Levin (2008: p. 135) show that the DO construction is associated with CP. Since give-roots inherently lexicalize CP, it overlaps with any CP meaning that might be associated with the DO construction in which they appear.

In contrast, the contribution of the DO construction (CP meaning) is revealed by the implementation of send-roots, according to Rappaport Hovav and Levin. Evidence for this is that whereas send-roots can select both a spatial target and a possessor in the PO construction, send-roots necessarily selects a possessor when implemented in the DO construction. The fact that send-roots exhibit variability between CM and CP meanings is evident from the comparison with CP-root give in the PO realization in (22) (examples taken from Rappaport Hovav and Levin (2008: p. 138)):

(22) a. I sent the package to Maria/London.
b. I gave the package to Maria/*London

Evidence that *send* is associated with CM meaning emerges from the acceptability of the spatial target London presented in 22a, which cannot occur with CP meaning of give in the same PO realization in 22b, since *give* necessarily selects a possessor, a requirement that London does not fulfill. Furthermore, Rappaport Hovav and Levin employ different testing procedures for the identification of CM meaning:

(23) a. Where did you send/*give the package?

b. I sent/*gave the package halfway/all the way to London

c. I sent/*gave the package to Maria, but she never received it.

The examples in (23) show that, unlike *give* (cf. chapter 4.2.4.1), it is only for *send* that a goal can be interrogated (23a) and a path modified (23b). In addition, the possibility of contextual cancelability of the send-action, but not of the give-action, also reveals an internal complexity of the path argument of the send-type (23c) (see also Oehrle, 1976; Wechsler, 1995).

Rappaport Hovav and Levin thus assume for PO realizations that while give-roots lexicalize change in possession but not change in location, send-roots lexicalize change in location but not change in possession. In this context, there is the option that the spatial goal of the send-root can be understood as a possessor if it is instantiated by an entity capable of possession.

However, Rappaport Hovav and Levin note that send-types have different properties in the DO realization than in the PO construction. This is shown (indirectly) by an analogous example with the CM root *throw* (2008: p. 144).

(24) a. Sam threw the ball to the first baseman/first base. (PO)

b. Sam threw the first baseman/*first base the ball. (DO)

In (18), Rappaport Hovav and Levin show that the first post-verbal object of the DO construction with *throw* must be occupied by an entity capable of possession. This suggests that while the PO construction (24a) can express both CM and CP, the DO construction (24b) must be associated with CP. The test procedure shown in (25), which Rappaport Hovav and Levin used to detect the CM component in (24b), further shows a complex path of throw in the DO construction, since it allows the transfer to be contextually cancelled:

(25) Sam threw/*gave the first baseman the ball, but he never received it.

Consistent with Rappaport Hovav and Levin's predictions, Beavers and Koontz-Garboden's (2020: pp. 107; see also Green, 1974: p. 103; Oehrle, 1976: p. 81) analysis shows that send-roots have the same semantic properties as throw-roots (cf. Rappaport Hovav & Levin, 2008: p. 134).

(26) a. #Mary sent London a letter. (DO)
b. Mary sent a letter to London. (PO)

Beavers and Koontz-Garboden note that the DO realization of *send* only occurs when London is understood metonymically to, e.g., London Office, an entity capable of possession. Understanding London as a purely spatial goal is incompatible with the CP meaning associated with the DO realization. Rappaport Hovav and Levin (2008: p. 138) also notice the metonymic London Office effect for the PO realization of *send*. Hence, both Beaver and Koontz-Garboden's and Rappaport Hovav and Levin's analysis suggest that the two realizations are associated with different event structures (CP with DO and CM with PO).

One might argue that if the DO realization were associated with CP, then each DO realization would have to involve an actual change of possession. This would be at odds with the possibility of cancelling the transfer in the case of throw- and send-roots in the DO construction. However, as Beavers and Koontz-Garboden (2020: p. 116) note, DO realization always involves *prospective* possession. This is justified by the fact that the target of the transfer must be occupied by, or reinterpreted as, an entity capable of possession (see also Gropen et al. 1989: 207). Moreover, the ability to possess is not restricted to animate entities, but also to inanimate entities that do not have to be reinterpreted as entities capable of possession, as the wall in *Kim gave the wall a fresh coat of paint*. Since, according to Beavers and Koontz-Garboden, these semantic properties follow from any DO realization, they are not only part of the root, but also of the DO construction independently from the root. Therefore, each of the presented analyses leads to the conclusion that the DO is a construction in itself, as per Construction Grammar.

4.2.4.3 Descriptive and schematic meaning components

Beavers and Koontz-Garboden's (2020) and Rappaport Hovav and Levin's (2008) analyses of the English dative alternation show that throw- and send-roots, which both lexicalize a spatial goal, additionally select a possessor through a metonymic London Office effect (cf. Section 4.2.4.3) in the DO realization. Beavers and Koontz-Garboden argue that the London Office effect shows that the DO construction associates a schematic meaning (in their words: templatic meaning). Conversely, the verb *give*, which only lexicalizes a possessor, does not select a spatial goal in PO realization. The fact that different root types show varying patterns of interaction with DO and PO realizations raises the question of what exactly the semantic contribution of roots and constructions is. Beavers and Koontz-Garboden present two basic answers for this: Either there is no overlap between the meaning of the verb and the construction, or there is in some cases. Based on the DO and PO realizations of *send*, (27) serves as evidence that there is no overlap (Beavers & Koontz-Garboden, 2017: p. 71; Beavers and Koontz-Garboden, 2020: pp. 108-9; Harley, 2003):

(27) $[_{v}P$ agent$[_{v'}$ $[_{v}$ v_{CAUSE} √send][PP recipient $[_{P'}$ P_{HAVE} theme]]]] (DO)
$[_{v}P$ agent$[_{v'}$ $[_{v}$ v_{CAUSE} √send][PP theme $[_{P'}$ P_{LOC} to recipient]]]] (PO)

In addition to the causative verbal head (v_{CAUSE}), the send-root (√send) specifies a manner of causation. The verbal head is understood as part of an underspecified schematic event structure CAUSE-HAVE (27a) or CAUSE-LOC (the latter is to be understood synonymously with CAUSE-BE-AT) (27b). While P_{HAVE} describes a result state involving a possessor, P_{LOC} describes only a spatial goal. These resultant states are components of the schematic meaning and, according to this analysis, not selected by the root, since *send* can occur with both meanings across constructions. Moreover, the resultant states apply across roots.

In contrast, when give-roots (*give*, *hand*, *pass*) are implemented in the structures in (20), the roots override both schematic meanings and level out the contrast between them.

That is, where the send-roots simply fill schematic meanings with content, the give-roots intrinsically specify schematic meanings by entailing CAUSE-HAVE as a resultant state in the PO realization. Thus, in agreement with Rappaport Hovav and Levin, Beavers and Koontz-Garboden claim that the give-roots themselves associate an event structure that is otherwise associated with the schematic construction, as shown when other root-types are implemented. *Send* is also telic in that it implies a prospective (though cancelable) BE-AT resultant state, while it additionally gets a CAUSE-HAVE meaning when implemented in the DO construction (cf. the London Office effect; Beavers & Koontz-Garboden, 2020: pp. 135f.).

Thus, according to Beavers and Koontz-Garboden, the division of meaning components contributed by the roots and the construction is not strictly drawn (against the assumptions of Harley (2003) and Pesetsky (1995)). Here, it is assumed that the idea that a root can either fill in a highly underspecified construction with a descriptive meaning component or override the construction with causative and resultative properties can be seen as consistent with Goldberg's (1995: 50f.) principle of verb-construction fusion (cf.

section 4.2.3; as highlighted by Beavers and Koontz-Garboden, 2020: p. 108).[10] Hence, the fact that the root can override the constructional meaning does not necessarily pose an issue for a modified version of Construction Grammar (cf., e.g., Welke, 2019). However, this topic still needs further investigation in future.

4.3 Cascades and Construction Grammar

4.3.1 Multilayered structures in action verbs

As we saw in chapter 2, cascades allow us to capture the structure of multilayered actions in ontology and cognition (cf. Goldman, 1970; Löbner, 2021). According to Löbner (2021), describing cognitive structures of actions as multilayered implies that the verbs

[10] This basic idea is also evident in the psycholinguistic evidence of Pickering and Branigan (1998), who showed experimentally that syntactic specifications of verb and construction can overlap (see also Pappert & Pechmann, 2014).

denoting these actions must be accordingly multilayered in their meanings. In fact, Löbner argues that verbs describing non-layered actions (i.e., basic actions) are among the exceptions of linguistic expressions. To illustrate this, Löbner (2021: p. 281) highlights the action verb *say*, among numerous examples. While one might assume that this verb is basic, Löbner argues that its meaning is constituted by a complex cascade. This consists of lower-level layers of actions such as 'articulating sounds', and a higher-level layer such as the co-occurring action ↥ 'producing words'. Hence, according to the analysis, the meaning representation of *say* is constituted by multiple actions coinciding in time and space.

Drawing parallels to Austin's (1962) speech act theory, Löbner also analyzes the cascade generated by the action verb *write*. The write-cascade includes, depending on the type of writing, the level 'writing by hand', which allows the higher cascade levels ↥ 'writing characters', ↥ 'writing text', and ↥ 'writing content' to be performed in parallel. The relation to speech act theory is that the writing-action can cascade from level of locution to level of illocution, but not to perlocution, since no change of state of any affected entities is accomplished by the action. While a causal relationship between a writing-action and another event is conceivable, e.g., ↥ 'writing a law' might lead to the implementation and enforcement of a law (i.e., the act 'writing a law' might cause someone to receive a fine), such causality is not accomplished by that writing-action in isolation, since a 'giving someone a fine' is a different action.

By contrast, in the case of causative action verbs such as *kill*, performing an action causes a result (*causative accomplishments*; Van Valin & LaPolla, 1997; see Dowty, 1979). In the context of a cascade, actions of this type by definition reach the level of perlocution.

According to Löbner (2019: p: 683), the event structure of *kill* can be paraphrased as 'x does something whereby x causes y to become dead'. More generally for causative verbs, their semantics can be schematized as follows (adopted from Löbner (2021: p. 288)).

(1) [do x, [predicate$_1$(x, (y))] CAUSE [BECOME predicate$_2$(x) or (y)]]

We read in (1) that the action from agent x ('do x') causes x or y to enter the state described by a variable that can be filled with a predicate (predicate$_1$). The action thereby causes a change of state of x or y. The verb *kill* therefore instantiates predicate$_1$ as 'x kills y' and predicate$_2$ as the result state 'x/y is dead'.

Löbner (2021: p. 288) shows that only slight modifications of this schematization are necessary to structure example (1) as a cascade. The level-generation in example (2) is based on the fact that (1) is to be divided into two components: First, the action denoted by the verb, and second, the predicate denoting the resultant state of the affected.

(2) [do x, [predicate$_1$(x, (y))] ↥ [x CAUSE [BECOME predicate$_2$(x) or (y)]]

In (2), the upward arrow (↥) indicates a level-generation by which the action denoted by predicate$_1$ cascades to a higher level denoted by predicate$_2$. Using the example of *kill*, this causal level-generation can be paraphrased as 'x does something that can be described by the predicate *kill*' ↥ 'x causes that y has the state 'dead''. Together, both predicates constitute a multilayered meaning of the verbal root *kill*.

In other roots, a resultant state layer is added by the specific syntactic realization. This comes about, for example, in *hammer flat* (Löbner 2020: p. 285). Where *hammer* is an atelic action verb, it obtains a telic meaning through the addition of the resultative state *flat* (telicity can be made evident by the modification in *She hammered the iron flat *for an hour*). Note that this meaning is added by the construction, since *hammer* occurs across constructions.

4.3.2 Level-generation from a Construction Grammar perspective

We turn our attention back to the examples (1) and (2) from chapter 4.3.1. It is assumed here that these schematizations can be reinterpreted in the context of Construction

Grammar. A basic principle in Construction Grammar is that a linguistic expression is an instantiation of an abstract type. According to the Lakoff-Goldbergian Construction Grammar analysis of argument structure constructions, this instantiation consists of the tokenization of schematic argument roles through verb-specific participant roles. For the DO realization of *give*, this means that the argument roles associated with the verb (giver, givee, given) must fuse with the argument roles of the DO construction (agent, recipient, theme).

We can draw parallels between principles of verb-construction fusions and cascade-theoretic notion of level-generation. According to principles of level-generation, the instantiation of an expression occurs by tokenizing an act-type and specifying a resultant state. This can be seen as analogous to instantiations of abstract construction types by concrete lexical items. Construction Grammar and cascade theory thus both capture different perspectives on the instantiation of actions and events. While the focus of cascade theory is which layers of an event structure can be instantiated (here, instantiation is assumed as part of a *mental model* (Johnson-Laird, 1983) or a broader *situation model* (Van Dijk & Kintsch, 1983) in working memory), the focus of Construction Grammar is on those aspects of an event structure that are relevant to grammatical and lexical processes of syntactic realization.

Cascade theory explains how the tokenization of an act-type asserts a level in a cascade, revealing what event layers underlie conceptualization of an expression. For example, we can assume that the DO realization of *send* generates a higher cascade level than the PO realization by not only generating a layer with a spatial goal, but also additionally generating a layer in which this spatial goal is a possessor (see chapter 4.2.4.2). The conceptualization of possession is presupposed by the conceptualization of a spatial goal at a lower cascade level (cf. (3)). In other words, a possession relation is induced by the DO realization of a verb that on its own only specifies that something is moved to a goal (e.g., *send*). Also, in the telic reading from the atelic verb *hammer* in *hammer flat*, we see that a higher cascade level is generated by the syntactic realization than the level that can be achieved by verb semantics alone (cf. (4)).

(3) A' x causes y to have z
↥
A x causes z to move to y by sending it

(4) A' x causes metal to become flat
↥
A x performs a hammering action

With respect to the comparison of cascade-theoretic and Construction Grammar perspectives on event representation, one might object that principles of Construction Grammar imply that the selection of the construction rather than the verb is crucial for the assertion of a cascade level. This is because the construction sets the interpretive frames for the implemented lexical material. For example, the addition of a resultant state *flat* in *Mary hammered the metal flat* would have to be seen as a result of instantiating a schematic resultative construction. That is, reaching a higher cascade level compared to *Mary hammered the metal* would be due to construction, not the verb. This could potentially go against principles of cascade theory, since it assumes that a specific cascade level is reached by tokenizations of actions, i.e., instantiations of concrete lexical item (and not by constructions denoting abstract action types). Since, according to Construction Grammar principles, the higher cascade level is not reached by the lexical contents (*hammer* and *flat*), but by the resultative construction ('X causes Y to obtain state Z'), it might contradict the assumptions cascade theory stating that lexical contents are responsible for more precise specifications and thus further generation of levels. The same objection could also apply to the assumption that if a cascade level with a possessor is achieved by the DO realization of *send* (cf. (4)) and not by the send-root itself, since it is the construction and not the verb that decides the asserted cascade level.

However, although the cascade level reached in these examples is due to the construction, it is still the lexical material that instantiates the abstract construction that defines the cascade level. Whether it is the construction or the verb that is the primary meaning component of the sentence is in principle irrelevant here (see the positions on this debate

presented in chapter 4.2.4). Ultimately, in both cases, the verb instantiates the construction and thereby receives its reading depending on the realization pattern (cf. Welke's (2019) Construction Grammar principle of compositionality). That is, whereas the construction can augment the valency of the verb, this augmentation does not occur without instantiation by the verb. For that reason, it is assumed here that the objection presented above does not form any contradiction between the principles of cascade theory and Construction Grammar. Thus, in this approach to Construction Grammar, one can argue that the lexical contents are independently responsible for generating the respective cascade level.

4.3.2.1 Send-roots

In (5) and (6), the send-root specifies an action that instantiates a more general causative action (DO x; agent x does something that can be by the predicate DO). The DO realization in (1) additionally introduces the embedded predicate HAVE, denoting CP meaning. On the other hand, the PO realization introduces the embedded predicate BE AT, associating the expression with CM meaning.

(5) [DO x, [send (x, y)] ↥ [x CAUSE [HAVE (y, z)]]

(6) [DO x, [send (x, y)] ↥ [x CAUSE [BE AT (y, z)]]

This analysis is based on assumptions by Beavers and Koontz-Garboden (2020) that the resultant state of *send* is decided by syntactic realization: The PO realization selects a spatial goal, while the DO realization instantiates a possessor as a spatial goal by the metonymic London Office effect (see chapter 4.2.4.3). That is, the send-root in itself defines a transitive action, with argument z introduced and semantically specified by the construction. Here, z specifies either a possessor goal in the DO realization or a purely spatial goal in the PO realization.

The upward arrow ↥ indicates a level-generation between the meaning components associated with each level introduced by the construction on the one hand and the verb

on the other. The instantiation of an abstract causative component 'DO x' associated with both the DO and the PO construction by the action 'x sends y' constitutes in itself a cascade level; this generates the higher level ↥ 'x causes y to have z' (DO) or ↥ 'x causes y to be at z' (PO).

As Beavers and Koontz-Garboden (2020: p. 136) note, however, send-roots on their own involve telicity, since the action described cannot be performed without bringing about a resultant state (or, rather, one cannot throw anything without bringing about some resultant state of the thrown entity). However, this state is specified only by the concrete realization. The embedded predicates CAUSE-BE AT or CAUSE-HAVE are thus introduced by the construction.

4.3.2.2 Give-roots

Give-roots (*give*, *hand*, *pass*), which subsume the schematic meaning of the construction, specify a CAUSE-HAVE event structure (i.e., CP meaning) in both the DO and PO realizations of *give*/*hand*/*pass*. (7) shows the event structure for both realizations:

(7) [DO x, [give/hand/pass (x, y, z)] ↥ [x CAUSE [HAVE (y, z)]]

Since, unlike send-roots, give-roots intrinsically specify a resultant state (cf. Beavers & Koontz-Garboden, 2020), give-roots leave no room for introducing additional constructional meaning, which levels out the meaning differences between the realization patterns of the dative alternation.

Moreover, the cascade-theoretic analysis explains why the actions of send-roots, but not give-roots, are contextually cancelable:

(8) Mary sent/*gave John the letter, but he never received it.

A cascade-theoretic analysis can explain why send-roots imply only a prospective, i.e., potential change of possession, regardless of their syntactic realization, whereas give-

roots necessarily imply a change of possession (cf. example (8)). This is because give-roots denote an action that takes place in time and space strictly parallel to the change of possession, whereas send-roots denote an action that must necessarily be completed and followed by an entity's motion to the final possessor in (possibly virtual) space. The cascade level generated by *Mary sent John the letter* must thus be at a lower level than the cascade level generated by *Mary* causes *John* to possess *the letter*. On the other hand, *Mary gave John the letter* generates a comparably higher cascade level, since the caused possession is entailed by the root. Thus, according to this analysis, the root *give* generates a higher cascade level than *send* (although lower cascade levels can still be traced back for give-actions). Although both roots can produce the same resultant state (*John* possesses *the letter*), only give-roots can imply this on their own. In other words, give-roots, unlike send-roots, necessarily generate the level of effectuation. It is noted that the examples (5) and (6) are cascadings to the level of effectuation by means of send-roots. However, in contrast to give-roots, this cascading only takes place under the here implicitly assumed conditions that the theme of the sending-action is, for example, not lost in a mailing system.

It is therefore the overlap of root and constructional meaning that accounts for the lack of possibility of contextual cancelability between the action and the result in give-roots. However, the delineation of meaning contributions between root and construction, which is overlapping in give-roots and discrete in send-roots, does not emerge in (7). Basically, it is clear in the present analysis that the level of effectuation must be reached only by give-roots, since only these roots include this resultant state as part of their lexical meaning (see Beavers and Koontz-Garboden, 2020: pp. 130ff. for an analysis of all types of so-called "true possession roots").

4.3.2.3 Creative use of roots

The cascade-theoretical perspective on constructional meaning constitution offers advantages in describing seemingly idiosyncratic examples. If we look at an example of the realization of an optional recipient argument (i.e., an argument not required by the

valency of the verb, or, in terms of Construction Grammar, a verb whose valency has been augmented by the construction), the prediction of constructional approaches (cf. Goldberg, 1995; Rappaport Hovav & Levin, 2008; Beavers & Koontz-Garboden, 2020) that CP meaning is associated with DO realization may face challenges:

(9) Die Mutter verbietet ihrem Kind den Computer. (DO)
'The mother forbids her child the computer.'

According to the constructional approach, the DO construction in (9) would have to coerce a CP reading of *forbid*. At first glance, the expression is difficult to understand as a type of a transfer event analogous to the implementation of send- and give-roots. However, on closer inspection, it becomes clearer how the sentence is associated with CP. A more basic schema can be associated from (9), in which one entity causes another entity to lose possession. This semantics thus matches the CP meaning except for negated possession: 'x causes y to not possess z'. This can therefore be modeled simply using a negation (¬) in (10).

(10) [DO x, [verbieten (x, y, z)] ↥ [x CAUSE [¬HAVE (y, z)]]

In addition, it can be seen in the context of a cascade that (10) is part of a larger cascade of actions that are performed in parallel by the action asserted in (10).

(11) A' *Die Mutter* causes *ihrem Kind* to not possess *den Computer*
↥
A Die Mutter verbietet ihrem Kind den Computer

In (11), we see a section of a cascade in which the expression of a prohibition leads to a subsequent resultant state involving that possession of something is lost. This cascade occurs only under the circumstance that social conventions are followed. However, these need not be followed, which is why the prospective loss of possession is contextually

cancelable in parallel with the send-root, as in *The mother forbids her child to use the computer, but the child continues playing anyways*. The division of different levels of the actions *forbid* and *send* therefore show that the execution of the respective action is clearly separable in time and space from the resultant state.

4.4 Finiteness

When we produce an utterance, we retrieve lexical units and combine them into coherent structures. These processes involve, in many languages (with a few exceptions, such as Chinese), combining an infinite component of the sentence with a finite one (Klein, 1994).

(1) The guests liked only red wine.

In sentence (1), we find the infinite sentence components *guests*, *red wine*, *like*, which together express a liking-situation and its involved participants. By making the verb finite (*liked*), a certain relation between these constituents is asserted. In many languages such as English, it is obligatory to mark the finite verb with tense (in this case simple past tense). According to conventional theory, the function of tense is to position the situation expressed by infinite sentence components relative to the time of utterance (before, during, or after). However, as Klein (1994) notes, this does not fully capture how the liking-situation relates to the time of utterance. Consider that it may still be a fact even during and after the time of utterance that guests only like red wine. Thus, it does not seem to be solely the fact that the guests only like red wine that is related to the time of utterance. To account for this observation, Klein suggests separating the time of the liking-situation – which encompasses the entire period in the lives of the guests in which they only like red wine – from the time at which it is asserted by the sentence that the liking-situation was the case. Klein calls the latter time the *topic time*. The assertion can be further modified and contextually constrained, such as *at yesterday's party*. This makes it clear that tense in (1) does not simply position a situation relative to time of utterance, but that it serves

to define a time when what is conveyed by the assertion was the case – e.g., at yesterday's party. Hence, this is potentially only a small part of the total time conveyed by the infinite sentence components.

On the basis of the topic time, we can identify the function of finiteness more precisely. As described above, tense is obligatorily introduced with finiteness and relates the topic time to the time of utterance. However, the English language has an additional category of morphosyntactic means of expressing temporal relations, namely verb aspect. According to Klein (see also Comrie, 1976; Moens & Steedman, 1988), aspect serves to highlight certain internal features of the lexical content of a verb. If we recall that the lexical content of event-denoting verbs consists of an initial and a final state (Klein, 1999; 2002; 2010; see chapter 3.2.1), an event structure with two stages can be identified in (2).

(2) The host had drunk all the wine.

What is conveyed by the infinite sentence components in (2) is an initial stage in which the wine has not yet been drunk and a final stage in which it has been drunk. We find evidence for this temporally complex event structure in the fact that the wine can only be drunk if it has not been drunk before (put in a different way: a wine bottle can only be emptied if it was not already empty; see chapter 2.3.1 on counterfactuality). What is asserted by finiteness in (2) (i.e., the past perfect tense) is that the drinking-event has come to an end, thereby highlighting a final time-interval of the event. If we consider a complementary example in (3), the opposite is the case:

(3) The host was drinking all the wine.

In sentence (3), finiteness is introduced through the progressive aspect, which presents the drinking-event as ongoing, by which its final stage has not been reached. English thus allows its speakers to express two different contrast relations in (2) and (3) on the basis of morphosyntactic aspect markings of the finite verb (progressive vs. perfect aspect).

Thus, according to Klein, tense and aspect differ in how they relate to the situation being described. Whereas tense asserts a holistic view on the event and relates the topic time to the time of utterance, aspect refers to an event-internal view by contrasting either the initial state with the final state of the event or vice versa.

4.4.1 Aspect and Construction Grammar

In the context of Construction Grammar, Michaelis (2004) claims that the function of progressive and perfective aspect in the English language (i.e., the progressive and perfective constructions) is to be a "stativizer". That is, when aspect morphology is combined with an event-denoting verb, it interacts with the verb in a way that coerces its denoted activity (in our words: two-stage event structure) into a static reading. The function of the progressive aspect is to maintain an understanding of the event stage in which the activity is still ongoing. Similarly to Klein (1994), this means that aspect operates on the verb semantics by asserting a contrast to the final stage.

In some cases, however, aspect may coerce the temporal structure of the verb semantics altogether. This is where the principles of Construction Grammar come into play (see chapter 4.2.3 for an overview of the principles of top-down compositionality). This becomes apparent when the progressive aspect is applied to verbs that already express a static meaning in themselves, such as *I am liking your explanation* or *He is remaining stable*. (Michaelis, 2004: p. 36). According to Michaelis, the progressive aspect in combination with a static verb *like* or *remain* coerces a different reading out of these verbs so that their static properties are replaced by an activity reading. This follows the principle of Construction Grammar that the constructional schema [*be* [verb stem]-*ing*] allows for the embedment of a verb lexeme, which is then assigned the basic meaning of the abstract construction. In this case, the constructional meaning coerces the verb lexeme to denote a dynamic activity, to which an ongoing or completed view can be applied.

4.4.2 Aspect and cascade theory

The idea that aspect indicates that a certain time interval of an event is under discussion can be related to the cascade theory (see chapter 2.2). Recall that cascade theory states that a human action involves several types of actions simultaneously: The action in (1) that a host drinks wine occurs in time and space under certain circumstances together with the more basic action that a host puts a glass to his or her mouth, that the host moves his or her arm, and that the host intends to perform each of these actions.

Conversely, it follows that the action of the host drinking all the wine also generates higher-level actions, namely that the host causes the wine to be drunk, that there is no wine left, that the host disappoints the guests, and so on. If we focus on the components of this action cascade central to the current context, it can be reduced to the two levels that an actor performs an action (the host drinks wine) and that by doing so, the actor causes something to happen (i.e., the host causes the wine to be drunk). If aspect serves to apply an ongoing view on an action and to contrast it with the final stage, i.e., the level of causation, it is straightforward to assume that aspect serves to assert a particular level in an action cascade (see also Löbner's (2020: 683ff.) analysis of cascade theory in the morphological domain). If we follow the Construction Grammar analysis of Michaelis (2004), aspect (and not lexical semantics; cf. the previous chapter) is the driving force behind the assertion of a cascade level. The present argument is based on the idea that aspect can separate actions from their induced changes by focusing either on the ongoing action as it unfolds (the progressive aspect) or on the resultant stage, which includes the consequences of the action (the perfect aspect).

5 A theoretical overview

In the theoretical part of this dissertation, two aspects of events and their involved human actions were examined. First, this concerned their multilayeredness. The act of paying for an item is an aspect of the act of acquiring that item (under the circumstance that the purchase is successful), and we can focus our attention on one of the two aspects or on both at the same time. Second, it concerned their temporal complexity. Our understanding of events and actions requires that we perceive changes over time and integrate them into coherent structures.

Chapter 2 laid out a foundation for how the layering of actions has been described in ontology (Goldman, 1970), and in chapter 3, it was explained how these ideas have been applied to cognition in the context of cascade theory (Löbner, 2021). A central insight concerns the logical relationship between levels of actions and events that do not appear to be causally related. Consider the following example: A leaf falls from a tree, and while falling, it rotates around itself. In this, an object (the leaf) exhibits two distinct events that coincide in time and space. Neither the falling nor the rotation of the leaf is caused by the other: The leaf does not fall because it rotates, and it does not rotate because it falls (but because of the wind resistance involved with falling). Crucially, however, the leaf would not rotate if it did not fall. This dependence between the falling-event and the rotation-event has been identified as a counterfactual type (following Lewis' (1973) definition of causal dependence). By recognizing counterfactual dependencies, the present analysis shows us that events that are apparently on the same level (Goldman, 1970) can be categorized in hierarchical taxonomies. In terms of cognitive representation, rotation is assumed to be part of falling (i.e., its rotation counterfactually implies its falling, although falling need not be brought to attention), whereas the opposite is not the case.

A more fine-grained perspective on the internal temporal and logical properties of events has been explored based on *Argument-Ttime Structure Theory* (Klein, 1999; 2002; 2010). If we adopt Klein's view of events, which overlaps in some ways with Altmann and Ekves' (2019) theory (*Intersecting Object Histories*) (with the fundamental

difference that Klein specifies events as abstract logical relations, whereas Altmann and Ekves specify events as the histories of multiple object state tokens and their intersection with other object tokens), then we can see that the view of logical constellations of time intervals with associated object states constitutes a distinct perspective on event representation. In addition, it provides a perspective on the relation between abstract event schemata and their constituting objects. This relation is evident in the sentence *The traffic light remained red*, which denotes an event unfolding over time that does not involve any object changes. This observation suggests that the perceptual system tracks objects over time on the basis of abstract logico-temporal schemata, and not just on the basis of perceived object changes. Hence, it suggests that events are mentally represented separately from their constituent objects, and that the basis of event representation is not necessarily constituted by, e.g., the instability of objects (see also Langacker, 1987). Together, the two aspects on temporal and multilayered structures of actions and events serve to create a more fine-grained view that goes beyond the conventional approaches such as the thematic roles approach (cf. Fillmore, 1968; Dowty, 1989, see Gerwien, Marberg & Nicolaisen, *in press*).

In chapter 4, it was examined how language relates to these cognitive structures. The presented analysis shows that verb semantics, constructional semantics, and finiteness can all be related to the idea that actions and events are temporally complex and multilayered. The following insights from grammatical and semantic theory were brought together with cascade theory: (i) The analyses of action verbs by Van Valin and LaPolla (1997) and Löbner (2021) show that abstract action schemas can be instantiated by lexical items that specify concrete actions and resultant states of affected entities. The principle that an abstract action schema is instantiated by the descriptive properties of a specific verb parallels the cascade-theoretic idea that a basic action can simultaneously encompass the tokenization of multiple action types. (ii) A similar principle is evident in theories of Construction Grammar (cf. Goldberg, 1995). In Construction Grammar, the construction is assumed to define a mental scene, for instance a causal relation between two or three entities, and the verb fills slots in the structure with descriptive properties, including the action type and referential properties of the entities involved.

Looking at a central grammatical operation, namely the assertion through finiteness, Klein shows that time plays a central role. More specifically, his analysis of time in language (Klein, 1994) shows that finiteness serves two purposes: First, tense relates the topic time of the expression relative to the time of utterance. Second, aspect serves to put an internal view on the described event by contrasting an initial state with a final state (progressive aspect), or vice versa (perfect aspect). In English, both aspects (progressive and perfect) are available to the speaker through the grammatical system and can be encoded on the verb. This is related to the proposed view of cascades in the following way: Because cascade theory is used to model multilayeredness of events, actions can be separated from the results they cause. This overlaps in a sense with Klein's view that aspect serves to distinguish between an ongoing action and the final states of the event entities involved.

In the following methodological chapter, some aspects of the abovementioned theoretical ideas will be examined experimentally. First, this concerns the dynamic nature of events and their intrinsic temporal complexity. Second, it concerns how language is mapped onto event representations. In the experimental part, we will leave the concepts of cascades and multilayered representations behind and focus on the hypothesis of the temporal complexity of events. However, for a description of the connection between the ideas of multilayered and temporally complex structures in cognition, the reader is referred to chapter 3.1.4.

6 Experimental hypotheses

We now turn our attention to selected theoretical aspects of event representation presented in the last part of the dissertation. First, the temporal complexity of event representations will be investigated. By using the visual-world paradigm, it will be investigated how aspect markings on double-object verbs modulate the comprehension of the associated transfer semantics of these verbs (Experiment 1). Next, we will look at a prediction that can be raised by principles central to Construction Grammar. This concerns the idea that if syntactic constructions associate event representations, then semantic information should be able to be integrated during language comprehension solely on the basis of structural processing of syntactic structures, i.e., sentences that do not contain lexical verbs. This will be tested examining the comprehension of double-object sentences containing novel verbs in the visual-world-paradigm (Experiment 2).

6.1 Aspect at the interface between linguistic and visual processing

As we have seen in chapter 3.2, following principles laid out by Klein (1999; 2002; 2010) and Altmann and Ekves (2019), it can be predicted that a mental representation of an event specifies a logico-temporal configuration of object states. Klein describes how multiple time intervals can be part of the lexical content of a sentence, while producing a finite expression involves accessing certain time intervals associating constellation of object states. For instance, when finiteness is accompanied by aspectual morphosyntax, constellations of either pre-states (progressive aspect) or post-states (perfect aspect) of the involved objects should be highlighted.

Hence, the English language allows its speakers to express different ways of viewing events in the world. In an event of transfer of possession, such as in the sentence *Mary has handed John the letter*, the perfect aspect of the verb (*has handed*) serves to indicate

that an action performed by the agent (*Mary*) has been completed, and at the same time, the recipient (*John*) has come into possession of the object (*the letter*). In contrast, the progressive aspect serves to assert an ongoing event stage, as in the sentence *Mary is handing John the letter*. Grammatical aspect thus allows speakers to express parts of a temporally complex lexical meaning by specifying that a particular event stage is under discussion together with the states of the associated people and objects (e.g. the recipient *John*'s states before and after he obtains possession of *the letter*) and the state of *the letter* as it changes location from one possessor to another.

The role of grammatical aspect has been investigated in the context of language processing. During the description of scenes depicting motion events (e.g., a car driving towards a town), i.e., in the context of language production, Stutterheim et al. (2012) compared eye-movements of speakers of different languages in which aspect is grammaticized (English) or not (German). The authors found that speakers directed their visual attention to areas in the scene that can be associated with ongoingness (e.g., the moving car) or endpoints (e.g., the town) in accordance with the existence of grammatical aspect in their respective language. The authors argued that grammatical features of the linguistic system influence how events are conceptualized, which is reflected in visual information uptake during pre-verbal conceptualization.

In the context of language comprehension, the influence of grammatical aspect in event comprehension has been investigated in the context of event segmentation (Feller, Eerland, Ferretti & Magliano, 2019), on the perception of event ongoingness and completedness (Madden & Zwaan, 2003; Magliano & Schleich, 2000; Anderson et al., 2008; Liao, Dijkstra & Zwaan, 2022), on mental simulation (Bergen & Wheeler, 2010; Glenberg & Kaschak, 2002) on event duration estimation (Flecken & Gerwien, 2013), on object state representation (Misersky, Slivac, Haagort & Flecken, 2021), on the activation of event roles such as locations (Feretti, Kutas & McRae, 2007) and instruments (Lombardi, Ford-Dominey & Ventre-Dominey, 2017), and on the salience of event participants in co-referential processing (Grüter, Takeda, Rhode & Schafer, 2018; Kehler, Kertz, Rohde, & Elman, 2008; Rhode, Kehler & Elman, 2006; Ferretti, Rhode, Kehler & Crutchley, 2009).

The latter studies on co-referential processing were concerned with how grammatical aspect manipulates the salience of agents and recipients when accessing mental models of transfer-of-possession events. In a sentence completion task, Rhode et al. (2006) showed that participants relied on grammatical aspect in sentences such as *John brought/was bringing a glass of water to Robert* to subsequently interpret an ambiguous pronoun that could refer to either the agent (*John*) or the recipient (*Robert*). Pronoun resolution was necessary in order to solve the task of completing a following sentence containing only the ambiguous pronoun. Results showed that participants were more likely to complete the following sentence by interpreting the pronoun as referring to the agent from the previous sentence after the past progressive aspect had been heard compared to the simple past tense. The authors argued that marking double-object verbs with the progressive aspect brings attention to the agent and away from the recipient (see also Madden & Ferretti, 2009).

In a study using the visual-world paradigm, Grüter et al. (2018) used similar linguistic stimuli as Rhode et al. (2006). The study showed that eye-movements reflect expectations of upcoming mentioning of agents and recipients in subsequent discourse. After hearing sentences such as *Donald brought/was bringing Melissa a fancy drink*, participants more often directed their visual attention to agents when the past progressive aspect (*was bringing*) was used than the simple past tense (*brought*), which reflects an increase in the likelihood of agents being involved in further activities described in subsequent discourse. Lee and Kaiser (2022) also manipulated aspect in transfer-of-possession events. In an online experiment using eye-tracking via webcam, the authors showed that eye-movements were reliably directed more towards the goal of transfer after the simple past tense (*Liam gave the ball to Paige*) compared to past progressive markers (*Liam was giving the ball to Paige*) in prepositional-object sentences were heard. Together, these studies show that the processing of grammatical aspect reliably directs attention to referents consistent with expectations of how aspect influences sentence meaning. This suggests that aspect modulates event representations during comprehension.

6.2 Structural processing

As we have seen in chapter 4.2, it is predicted by constructional approaches to language (e.g., Construction Grammar) that event structure is associated not only with lexical verbs, but additionally with basic sentence forms. According to this view, the DO construction is associated with a basic transfer-of-possession event. Applied to a psycholinguistic context, this challenges a prevalent assumption that event structure is only retrieved by decoding the lexical content of a verb, by which the verb is considered to be the main determiner of sentence meaning. This verb-centered view follows the principles of lexicalist accounts of sentence comprehension (Juliano & Tanenhaus, 1994; MacDonald, Pearlmutter, & Seidenberg, 1994). Instead, hypotheses of constructional approaches go in line with structural approaches to sentence processing, according to which the generation of syntactic structures and the inference of sentence meanings is dissociated from lexical access.

In a study supporting the structural approach to sentence processing, Thothathiri and Snedeker (2008) conducted structural priming experiments with the dative alternation. Structural priming refers to a tendency primarily observed in language production of speakers re-using abstract syntactic representations available from recent processing (Bock, 1986; Konopka & Bock, 2009; Bock & Loebell, 1990). In their study investigating three- and four-year old children's interpretations of ditransitive utterances, Thothathiri and Snedeker (2008) compared the influence of double-object and prepositional-object sentence primes on subsequent comprehension of temporarily ambiguous double-object and prepositional-object sentence targets (*Show the **hor**se the book*; *Show the **hor**n to the dog*). Eye movements recorded during target descriptions showed that double-object sentences, which encode the recipient as the first post-verbal argument, primed the expectation that the first post-verbal object in a new target sentence would also be a recipient. Importantly, the effect was observed both between primes and targets containing the same verb (*give*) as well as different verbs (e.g., *show*, *bring*). Based on this, the authors argued that listeners rely on abstract structural information during sentence comprehension.

In line with this finding, Ziegler, Snedeker and Wittenberg (2018) showed that structural priming is not only sensitive to linearizations of thematic roles associated with a verb (such as the ordering of arguments in the *Mary gave John a book* vs. *Mary gave a book to John*), but to more fine-grained event structural differences associated with different syntactic realizations. They contended that if thematic roles are primed, then different syntactic constructions involving the patient of a light verb (such as "her son" in "The mother is giving her son a hug" or "The mother gives a hug to her son") should have a stronger influence on the choice of syntax in descriptions of transitive events (e.g., a depiction of a mother giving her son a kiss on the cheek) compared to the recipient of a ditransitive verb (like "her son" in "The mother is giving her son an apple" or "The mother gives an apple to her son"). Since the study indicated no discernible distinction in the priming effects between light and ditransitive verbs, the researchers concluded that priming did not exhibit sensitivity to varying syntactic expressions of the patient of light verbs. Instead, the effect was attributed to syntax and its associated event structure, which, in the case of light verbs, was presumed to be 'X acts on Y' across different syntactic representations. As per the authors' interpretation, this suggests that event structures, rather than arrays of thematic roles, are associated with syntactic forms.

The above-mentioned studies show evidence that not only syntactic, but in addition semantic representations are re-used in subsequent language processing. Although studies typically localize structural priming effects at a syntactic level (e.g., Branigan, 2007; Branigan & Pickering, 2017; Pickering & Branigan, 1998), several further priming studies have shown that the semantic level also plays a role (e.g., Cai et al., 2012; Pappert & Pechmann, 2014; Ziegler & Snedeker, 2018). The different syntactic and semantic foci of these studies relate to two main hypotheses about what forms the basis of structural encoding during language production (i.e., the mapping semantic argument structures onto syntactic sentence forms): either it is assumed that mapping from conceptual, preverbal representations to linguistic forms goes via the verb, or it is assumed that concepts can be mapped directly onto linguistic structures bypassing any restrictions of the verb (see Baumann, Pappert & Pechmann, 2021).

In a comprehension study by Kaschak and Glenberg (2000) supporting an interlinked view of syntactic and semantic representation, the authors tested whether participants could rely on argument structure configurations to derive sentence meanings in the absence of the lexical verbs licensing them. Their study shows that after reading DO sentences containing novel, denominal verbs, i.e., non-conventional verbs that had been derived from nouns such as *a crutch* as in *Lyn crutched Tom her apple so he wouldn't starve*, participants chose the interpretation 'To transfer using a crutch' more often than 'To act on using a crutch', despite of the fact that both interpretations could apply to the meaning of the DO construction. The authors interpreted this effect as evidence for that listeners are able to derive rudimentary event structures from syntactic parsing independently from the semantics of the lexical verb (see also Bencini & Goldberg, 2000; Ziegler, Bencini, Goldberg & Snedeker, 2019). However, it cannot be ruled out that subsequently reading these potentially leading answers would cause backward priming of the DO constructions. Hence, the study does not say whether participants would retrieve a semantic representation during online-processing or simply revise their interpretation according to the options given.

Since Kaschak and Glenberg's (2000) study did not investigate online processing of DO constructions, the study does not tell us how novel verbs were interpreted as they were presented to the listener. However, there are accounts setting out to explain the involved interpretive mechanisms (specifically for denominal verbs, see Hsiao and Michaelis (2021); see also principles of *co-composition* in Pustejovsky (1995) and *conceptual blending theory* in Fauconnier and Turner (2002)). What these different accounts have in common is the assumption that a lexical item is accommodated to the context they appear in. For novel (denominal) verbs, this means that the argument structure they appear in imposes semantic restrictions on their interpretation. During incremental processing of the DO construction containing a novel verb, this idea predicts that the first post-verbal object of the sentence will induce a causative interpretation of the sentence, and that this interpretation will be modified to a caused possession interpretation upon integration of the second post-verbal object. Whereas research in prediction in language comprehension has investigated how the semantic selectional restrictions of the verb triggers anticipation

of referents that fulfill these restrictions prior to their mentioning, this implies conversely implies that the argument structure in which novel verbs appear will restrict the interpretation of the verb. In the following chapter, we will look at research that has investigated the incremental processing of sentences.

6.3 Prediction in event comprehension

Studies in event comprehension have investigated predictive mechanisms to examine which information contained in a sentence is used during online processing. In a classic study by Altmann and Kamide (1999), the visual-world paradigm was used to uncover how sentence elements guides eye-movements, reflecting the online interpretation of the sentence. The results shows that while hearing *The boy will eat…*, participants were more likely to fixate a concurrently visible edible referent (e.g., a cake) than a non-edible one (e.g., a toy car) after verb onset and before onset of the theme (e.g., …*the cake.*). The authors argued that verb semantics are used to predict plausible referents during syntactic parsing. In a further study using a similar experimental design, Altmann and Kamide (2009) raised the question of whether, in addition to predicting upcoming referents, listeners also predict the different states of these referents. Participants were presented with scenes containing, e.g., a woman, a wine glass, a bottle of wine, and a table. The scene then disappeared, leaving a blank screen, and participants heard a sentence defining a context in which an object (the wine glass) either had been moved or not, such as *The woman will put the glass on the table* or *The woman is too lazy to put the glass on the table*. This was followed by a target sentence *She will then pour the wine into the glass* describing an interaction between the agent (*The woman*) and the theme (*the glass*) from the context sentence. The results show that after hearing a context sentence describing that the object had been moved to a new position (e.g., the table), and while hearing a target sentence defining an interaction with the moved object, participants were more likely to fixate the area of the blank screen where this new position of the object had previously been shown, compared to its original position in the scene. This shows that mental representations generated through the linguistic input were mapped onto

representations of previously seen objects. Importantly, fixations on the new moved position already increased significantly after verb offset (*pour*). Together, these studies on predictive mechanisms show evidence that during incremental sentence comprehension (i.e., as each of the sentence arguments are auditorily presented), referents and their states (which in this case are understood as locations) are predicted prior to their mentioning.

According to Zacks et al. (2007), this prediction process is assumed to rely on a representation, a so-called *working model*, which is temporarily maintained in working memory and continuously updated as the perceptual input interacts with event schemata retrieved from long-term-memory. This is achieved by associating the current perceptual experience (auditory linguistic input or visually perceived input) with a generalized event schema retrieved from long-term memory. This creates an intermediate representation, a working model, which enables us to understand the world as it unfolds around us and, based on earlier experiences, to predict what will usually happen next in a given situation (see also Zacks & Tversky, 2001; Radvansky & Zacks, 2014).

6.4 The present experimental study

So far, we have seen evidence that (i) past progressive aspect markings on DO verbs (compared with simple past tense markings) increases the likelihood of expecting an agent (versus a recipient) to be re-mentioned in subsequent discourse (Grüter et al. (2018); (ii) that structural priming has effects on the order of thematic roles (Pappert & Pechmann, 2014; see Baumann, Pappert & Pechmann, 2021) and on an event structural level (Ziegler et al., 2018), suggesting that argument structures can be primed independently of the verbs that typically select them; (iii) that readers in offline-studies (sentence meaning choice tasks) are able to infer transfer semantics from the syntactic structure of the DO construction independently of verb semantics (Kaschak & Glenberg, 2000); (iv) that event comprehension involves prediction of referents and their states (i.e., locations) (Altmann & Kamide, 1999; 2009).

However, it is still an open question how listeners use morphosyntactic features of an auditory linguistic input during real-time sentence processing. More specifically, it is not

clear whether listeners use aspect as a cue to predict and integrate referents and their states (i.e., 'before'- and 'after'-states) while a sentence unfolds and after a final interpretation is available. Furthermore, it is unclear whether listeners use the syntactic structure of DO constructions containing denominal verbs to derive transfer semantics immediately after sentence end.

In two experiments, the following questions were asked: (1) Do listeners use the verb aspect to predict and integrate 'before'- and 'after'-states of referents during real-time processing? (Experiment 1) (2) Can the retrieval of sentence meaning, on which the integration of specific states of referents relies, be based on morphosyntactic processing alone (i.e., by processing syntactic structure without a lexical verb)? (Experiment 2).

To answer these questions, a modified visual-world paradigm (Cooper, 1974; Tanenhaus, Spivey-Knowlton, 1995) was used to uncover whether subtle aspectual distinctions of descriptions of sentences conveying transfer of possession can be used to guide overt visual attention in static visual scenes (containing two humans and a movable object) to object locations that can be associated with the 'before'- and 'after'-states of event entities (i.e., before and after an object was transferred between an initial and a final possessor).

In both experiments, participants viewed scenes containing two human characters and a movable object and heard two types of sentences describing an action involving these visual referents. In Experiment 1, the sentence type was a DO construction containing a conventional DO verb marked with aspect, e.g., *Mr. Pink is handing/has handed Mr. Green the sock*. In Experiment 2, the sentence type was a DO construction containing a novel, denominal verb marked with aspect, e.g., *Mr. Pink is clothespinning/has clothespinned Mr. Green the sock*.

Aspect marking was manipulated to test whether listeners predict and integrate referent states when processing conventional DO verbs (Experiment 1), and to investigate whether lexical and structural processing mechanisms can be distinguished during the comprehension of DO constructions containing novel DO verbs (Experiment 2). Hence, the experiments aimed to determine whether and how quickly abstract morphosyntactic features of linguistic input (aspectual morphosyntax and syntactic structure) can contribute to the mapping of comprehended language onto visual scenes.

6.4.1 Experiment 1

In Experiment 1, we investigated whether auditory comprehension of DO constructions marked with grammatical aspect (progressive vs. perfect) can direct visual attention to areas in a visual scene that can be associated with a 'before' and an 'after' a recipient has obtained possession of a transferred object. Figure 1 shows a visual scene that was presented to participants. Concurrently, participants heard a sentence describing that one human character either is handing or has handed a previously introduced object to the other human character.

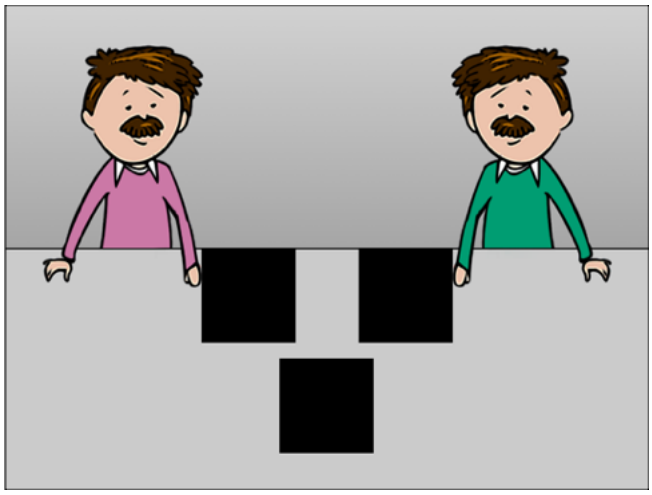

Figure 1. Example visual stimulus used in both experiments of a scene that was shown during and shortly after the auditory presentation of a double-object sentence containing a conventional verb (Experiment 1) or a novel, denominal verb (Experiment 2) marked with grammatical aspect.

Of main interest was visual attention to masked locations next to humans corresponding to agents and receivers of the auditory input (see Figure 1). Participants were expected to interpret the masks as indicating a possessive relation between one of the humans and a previously introduced object (e.g., a sock) (see the Materials chapters for more information on trial structure and stimulus material).

Eye-movements were measured while participants performed a sentence comprehension task. Participants first freely inspected a visual scene that first showed two human

cartoon characters and while hearing an auditive sentence introducing these visual referents, e.g. *This is Mr. Pink and Mr. Green*. Then they saw a scene containing only two objects (e.g., a clothespin and a sock), and they heard an auditive sentence introducing these, e.g., *This is a clothespin and a sock*. Finally, they saw a scene containing two human characters and three masked locations, to which the objects could have been moved while hearing descriptions such as *Mr. Pink is handing Mr. Green the sock* (progressive condition) or *Mr. Pink has handed Mr. Green the sock* (perfect condition) (see Figure 2). During the auditory presentation of these descriptions, the three masks covered the object and two other potentially new locations of the object next to each human character. After sentence offset, the participants' task was to indicate which mask was most likely to cover the moved object. Since the visual stimulus gave no clue as to the final location of the object, participants had to rely on the auditory description to infer the possible location of the object.

Of interest was whether eye-movements would show that participants relied on grammatical aspect to make inferences about the location of the object, reflecting modulation of their mental representation of the event as ongoing or completed. Specifically, we expected that in the progressive condition, participants would fixate agents and agent masks more than recipients and recipient masks, as the auditory sentence would describe the activity of an agent as ongoing and still in possession of the object. In the perfect condition, we expected participants to fixate recipients and recipient masks more than agents and agent masks compared to the progressive condition, as the auditory sentence would describe the transfer as completed (by which a recipient had obtained possession of the transferred object).

6.4.2 Experiment 2

In Experiment 2, it was investigated whether similar effects of verb aspect as in Experiment 1 could be observed when the DO verb (e.g., *hand*) in the auditory sentences was replaced by a novel, denominal verb marked with grammatical aspect, such as *Mr. Pink is clothespinning Mr. Green the sock* (progressive condition) or *Mr. Pink has*

clothespinning Mr. Green the sock (perfect condition) (for a description of the choice of linguistic material, see chapter 8.2.) We were interested in whether participants could infer the transfer meaning of without accessing a lexical representation of the sentence verbs, i.e. by relying on the syntactic structure of the sentence alone.

7 Experiment 1

In this experiment, grammatical aspect markings (progressive vs. perfect) on DO verbs (*hand* and *pass*) were used as manipulation to see whether listeners associate areas in a visual scene with a 'before' and an 'after' a visual human character corresponding to a recipient of an auditory description has obtained possession of a transferred object. It was expected that after the progressive aspect is heard, more visual attention would be directed to a mask next to an agent (indicating the agent's possession of the transferred object), and after the perfect aspect was heard, it was expected that visual attention would be directed more towards a mask next to a recipient, indicating that the recipient has come into possession of the object. This would suggest that the described event is understood as completed, since the final stage of a transfer event is defined by a recipient having come into possession of an object.

7.1 Participants

20 native speakers of English participated. The participants were students aged 18-35 recruited from the University of Aberdeen, the Robert-Gordon University and Heidelberg University. Prior to the experiment, participants filled out a language questionnaire in which they reported details of their linguistic background. Only data from participants who were native speakers of English and who had no exposure to languages other than English before the age of 5 are reported. Each participant received a compensation of £4 or course credit.

7.2 Materials and design

The structure of each experimental trial (20 experimental trials; 40 non-transferable filler trials; see Appendix A) consisted of the initial visual presentation of images created in GIMP at 1024*768 resolution of either two female or two male human cartoon characters.

A visual scene contained two human characters that were visually identical and could only be distinguished from each other by the color of their clothing. This was done to reduce the variation in visually salient features in the human characters' appearance. To further avoid distracting labels, genders and colors were used as a way to refer to them, e.g. *Ms. Pink*/*Mr. Pink*. The colors were chosen from a colorblindness friendly palette provided by www.cookbook-r.com. The referent types and their colors were balanced across all experiments. In addition to images presenting the human characters, images each presenting two objects (e.g., a clothespin and a sock) were designed.

In the trials (n = 20), participants heard a critical sentence containing a DO verb marked with either the present progressive aspect (e.g., *Ms. Pink is handing Ms. Green the cookie*; n = 10 trials) or present perfect aspect (*Ms. Yellow has handed Ms. Blue the ice cubes*; n = 10 trials) (see 3. in Figure 2). After the presentation of the critical sentence (average sentence duration in the progressive condition were 3321 ms, and 3391 ms in the perfect condition), the display remained visible for an additional 2000 ms. 100 ms after the 2000 ms, the letter T, which was either tilted to the left or the right, was presented on each of the three masks (see Goller et al., 2020). This was implemented in order to motivate eye-movements to regions that would be relevant for the analysis and to prevent strategies of systematically fixating the middle of the screen and only processing visual information peripherally. Hence, the recognition of tilt of the T demanded foveal attention, i.e., direct fixations on the mask. The tilt of the T varied on each mask from trial to trial.

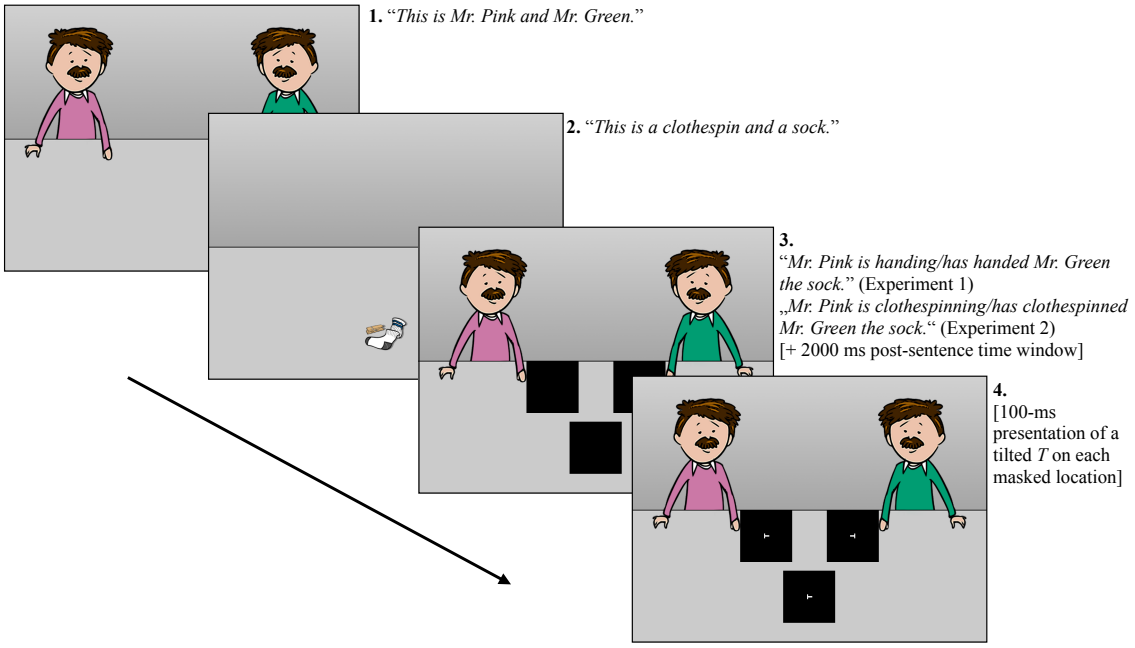

Figure 2. Example trial structure used in both experiments. The same visual scenes were used for both critical items and non-transfer filler items. First, participants saw two human referents that were concurrently auditorily introduced. Second, they saw two objects that were auditorily introduced. Third, during the presentation of the DO construction, the human referents reappeared with three masks covering potential locations of the previously introduced objects. At the same time, in the experimental trials, participants heard a DO construction containing either a conventional verb (Experiment 1) or a novel, denominal verb (Experiment 2) marked with aspect. The sentence presentation was followed by a 2000 ms time window and a 100 ms presentation of a tilted T on each mask.

The filler trials (n = 40) followed the basic structure of the experimental trials. However, transfer descriptions (see 3. in Figure 2) were replaced by auditory sentences describing either a transitive human-object interaction (n = 20), e.g., *Ms. Green is lifting up the pot*, or a transitive human-human interaction (n = 20), e.g., *Ms. Yellow is talking to Ms. Blue*. The fillers categories were designed to include all three visual referents in ways that did not restrict the meaningful interactions between objects and humans to only the experimental stimuli. The sentences presented in filler trials also used progressive and perfect morphosyntax. The position of the agent of auditive descriptions in the visual scene (left or right) was counterbalanced across gender, referent type, trial type and the conditions (progressive and perfect). The overall design of experimental trials and filler trials only differed in the type of experimental sentence (transitive vs. ditransitive).

All auditive stimuli were recorded in a studio with a native speaker of British English. The sentential stimuli were recorded uniformly and in their entirety. The experiment was programmed and compiled in SR Experiment Builder.

7.3 Task

Participants were instructed to listen to the description and decide which of the three masks covered the object. After a decision time of 2000 ms after the sentence, the letter T (tilted either left or right) appeared on each mask for 100 ms. Participants' task was to press the left or right button (a and l on the keyboard) to report the direction of the T they saw on the mask they assumed covered the object.

7.4 Procedure

Participants were seated at an EyeLink 1000 Plus eye tracker (1000 Hz sampling rate) and read instructions on a screen, which were paraphrased by the experimenter to ensure participants understood the task. Participants were also familiarized with the experiment by seeing four filler trials. Apart from the instructions and practice phase, each experiment lasted approximately 18 minutes.

Participants started each trial by pressing a button. In each trial, the referents were introduced with a recorded description such as *This is Mr. Pink and Mr. Green*. This was done to familiarize participants with the human referents and their positions. Participants then saw a visual scene containing two movable objects, e.g. a clothespin and a sock, and heard a similar *This is a clothespin and a sock* (for a more detailed description of the choice of objects, see chapter 8.2.). Each trial contained a different pair of objects. Then, the human referents reappeared, along with three black square masks covering the position where the objects had previously been seen (the base mask) and two additional positions next to the human referents (masks next to the human characters referred to as agents and recipients in the auditory descriptions in the experimental trials) (see Figure 2).

To motivate auditory attention to the content of the descriptions, participants received two kinds of comprehension questions in a pre-defined subset of filler trials (n = 8). The first type of question came after fillers showing human-object interactions and asked which of the human referents performed the described action, e.g., *Who picked up the pot?*. The second question type came at the end of filler trials that involved interactions between human characters and asked participants to identify the patient of the scene, e.g. *Who was spoken to?*. Participants were asked to answer whether the question applied to the left or right human character in the visual scene by pressing a (left) and l (right) on the keyboard. By reporting the position of the referent as opposed to the identity of the referent (e.g., Ms. Yellow), participants could not rely on the phonological buffer but had to map the linguistic input onto the visual scene. Furthermore, in different pre-selected filler trials (n = 8), the masks disappeared 2000 ms after the sentence end, revealing the object in one of the three possible locations. Afterwards, participants reported whether they had expected the object to be in that particular location by pressing 'yes' or 'no' (a and l on the keyboard). This was not intended as a comprehension task, but simply implemented to make it clear to the participants that the location of the object could vary between masks.

7.5 Results

7.5.1 Eye-movement data pre-processing

Under the assumption that linguistic input can drive eye-movements that reflect online prediction and interpretation (the *linking hypothesis*, Just & Carpenter, 1980; Magnuson, 2019), the present interest was to evaluate how many fixations were recorded in the different critical regions of the display (human referents and masks) over time: (a) during the verb region, (b) the recipient presentation, (c) the theme presentation, and (d) the post-sentence region of 2000 ms. There were several reasons for examining the number of fixations as opposed to, for example, fixation durations. First, since the analysis includes

the study of anticipatory eye-movements during the unfolding of the sentence, it was hypothesized that an auditory stimulus could elicit a fixation to a particular AOI. By analyzing the number of fixations to an AOI within an interest period (IP) instead of the time the eye rested within that AOI, it was expected to capture an aspect of eye-movements that, in terms of covert attention, more closely reflects the dependence between understanding a word and making an eye-movement. Second, if a fixation fell within an AOI, it could remain there during the following IP, potentially adding data to the fixation durations during that IP that were not of interest in the analysis. By instead analyzing the number of fixations within an AOI that occurred after a particular IP started, it was expected to reduce such carry-over effects across IPs.

The time course of each sentence was divided into three individual interest periods (IP) (IPa = verb, IPb = recipient, IPc = theme, IPd = post-sentence). Each time interval was defined by the acoustic onset and offset of the critical words/phrases in each sentence. Note that the duration of each of the resulting time intervals was slightly different for each sentence because the phonetic durations of the individual words defining each IP were different across topics. For each time interval, all fixations that started before the beginning of the respective word/phrase defining the interval or that started after the offset were removed. This was done because only fixations that start after the onset of the respective word can be confidently interpreted as potentially driven by the respective part of the auditory stimulus.

Of interest were also potential differences with respect to the interpretation of the sentences, which were assumed to be reflected in the eye-movement behavior in the time interval between the sentence offset and 2000 ms thereafter (the post-sentence period of interest). Again, it was assumed that only fixations recorded after the start of the post-sentence period of interest could be safely interpreted to reflect any interpretive effects that may result from the integration of all words in a sentence. All fixations that started before the offset of the theme referent were removed. As it is commonly believed that there is a delay between understanding an auditory input and this eliciting an eye-movement of approximately 200 ms (Altmann & Kamide, 2004; Matin et al., 1993), eye-

movement data, specifically the onset and offset of each fixation, were shifted ‘towards the right’, i.e., by adding 200 ms to fixation onset and fixation offset.

7.5.2 Interest areas

Five areas of interest (AOIs) were defined per display: agent, recipient, agent mask, recipient mask and base mask. Out of these, we consider only the first four as critical. Of interest was whether the comparison of the sum of fixations in a specific critical AOI differed when comparing between conditions, given that the sum of fixations differed from trial to trial and from subject to subject. A fixation that landed in a specific critical AOI was interpreted as ‘success’ (1), and a fixation that did not land in a specific AOI as ‘failure’ (0). The focus was on the analysis of successes. Thus, for each subject, successes were aggregated over all trials from each condition to derive the total number of fixations (sum of ‘trials’) and the sum of fixations in a specific critical AOI (sum of ‘successes’). For example, for subject 1, there were a total of 23 fixations across all 10 trials from the progressive (PROG) condition, and 14 of these were recorded in the agent’s AOI (IP verb). Successes and attempts were calculated separately for the two conditions for all participants and all AOIs.

7.5.3 Visual inspection of the data

Figure 3 shows the relative proportions of fixation numbers in both conditions for all AOIs in the four interest periods. The ‘verb’ panel suggests almost no differences between conditions during the first interest period (verb onset to recipient onset). The ‘recipient’ panel, which shows the distribution of fixations during the second interest period (recipient onset to theme onset), suggests potential differences between conditions in the agent and recipient AOI, as well as in the agent mask AOI. The ‘theme’ panel, which shows the distribution of fixations during the third interest period (theme onset to theme offset), suggests more fixations to the agent AOI and the agent mask AOI in the progressive condition, and more fixations to the recipient AOI. The ‘post sentence’ panel, visualizing the

distribution of fixations in the post-sentence period, suggests more fixations on the agent mask in the progressive condition and more looks to the recipient mask in the perfect condition.

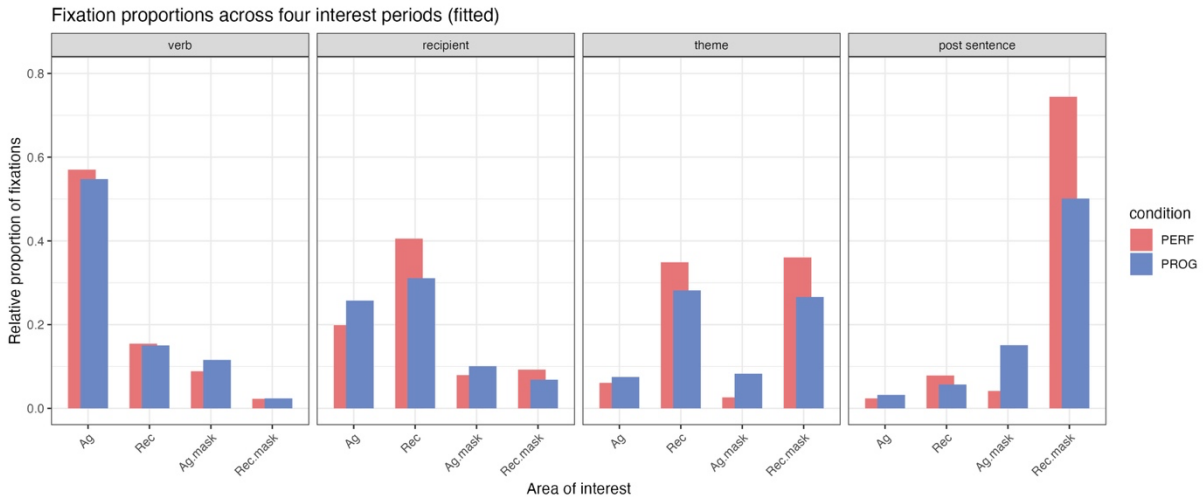

Figure 3. Experiment 1. Relative proportions of fixation in four interest periods (verb, recipient, theme, post-sentence) per critical AOI and condition (aspect); fitted values.

7.5.4 Modeling

R (version 2022.07.1+554) and the brms package (Bürkner, 2017; 2018) were used to set up 16 separate models to compare fixation data between conditions for the 4 critical AOIs (agent, recipient, agent mask, recipient mask) in three critical interest periods (recipient onset-offset, theme onset-offset, post sentence onset-offset). Model specifications were almost identical for all models. An intercept for the condition term was specified (formula: n | trials(total) ~ 1 + aspect, where 'n' is the sum of fixations in the respective AOI aggregated over all trials from a condition per subject, and 'total' is the total number of fixations aggregated over all trials from a condition per subject; aspect was dummy-coded: 0 = PROG, 1 = PERF). All models allowed slopes to vary for the random variable 'subject' for both aspect conditions, thereby accommodating random variance induced by inter-individual differences. By adding 'binomial(link = logit)' as the value for the 'family' function, it was specified that estimations were calculated in the logit space. Priors

for the condition intercepts were set as normal (mean = 0.00, SD = 50). Standard deviations of random effects had Student's t distributed priors, $b0 \sim t_{(3, 0, 2.5)}$. All models were fitted with four Markov Chain Monte Carlo (MCMC) chains, initially each with 12,000 iterations, 2,000 warmup samples (thin = 1). This resulted in a total 10,000 posterior samples per chain. These were combined into one posterior sample consisting of 40,000 samples. Convergence and stability of the Bayesian sampling for each model was assessed using R-hat, which should be below 1.01 (Vehtari et al., 2019), and Effective Sample Size (ESS), which should be greater than 1,000 (Bürkner, 2017). In a few cases, initial specifications led to convergence problems. If this was the case, the number of warm-up samples was adjusted and/or the 'adapt_delta' argument was increased, as recommended by Bürkner (2017). The brms hypothesis function was used to compare the differences between conditions as derived from the posterior distributions (always testing theoretically derived hypotheses for each AOI). Table 1 reports group differences on the probability scale (applying the brms internal inv_logit_scaled-function on the posteriors) and the corresponding 95% credible intervals (CI). Given the model specifications and given the data, a difference was considered as meaningful if the difference is estimated to be different from 0, and 0 does not fall within the 95% CI. In addition, the Bayes factor was provided for the hypotheses that were tested.

Table 1. Experiment 1. A star in the right-most column indicates at least a meaningful difference between conditions, given the data and the model specifications. The higher the Bayes factor the more evidence for the tested hypotheses.

Interest period	Hypothesis tested	Comparison of AOI	Estimated intercepts (mean with 95% CI in [])	Estimated mean difference with 95% CI in []	Bayes factor	
Verb	PROG > PERF	Agent	PROG: 0.54 [0.45, 0.64] PERF: 0.57 [0.39, 0.72]	0.02 [-0.04, 0.09]		
	PROG < PERF	Recipient	PROG: 0.15 [0.10, 0.21] PERF: 0.35 [0.07, 0.29]	0.00 [-0.04, 0.05]		
	PROG > PERF	Agent mask	PROG: 0.11 [0.06, 0.18] PERF: 0.08 [0.03, 0.22]	-0.03 [-0.07, 0.01]		
	PROG < PERF	Recipient mask	not modeled due to insufficient data			
Recipient	PROG > PERF	Agent	PROG: 0.26 [0.17, 0.37] PERF: 0.20 [0.08, 0.37]	-0.06 [-0.11, -0.01]	29.77	*
	PROG < PERF	Recipient	PROG: 0.31 [0.22, 0.40] PERF: 0.40 [0.23, 0.59]	0.09 [0.03, 0.16]	97.11	*
	PROG > PERF	Agent mask	PROG: 0.10 [0.04, 0.19] PERF: 0.08 [0.01, 0.32]	-0.02 [-0.09, 0.04]		
	PROG < PERF	Recipient mask	PROG: 0.09 [0.02, 0.30] PERF: 0.07 [0.03, 0.13]	0.02 [-0.02, 0.06]		
Theme	PROG > PERF	Agent	PROG: 0.08 [0.03, 0.16] PERF: 0.06 [0.01, 0.03]	-0.02 [-0.07, 0.03]		
	PROG < PERF	Recipient	PROG: 0.28 [0.22, 0.34]	0.07 [0.00, 0.13]	19.58	*

			PERF: 0.35 [0.21, 0.50]			
	PROG > PERF	Agent mask	PROG: 0.08 [0.03, 0.17] PERF: 0.02 [0.00, 0.16]	-0.06 [-0.12, -0.01]	50.95	*
	PROG < PERF	Recipient mask	PROG: 0.26 [0.16, 0.40] PERF: 0.36 [0.15, 0.64]	0.09 [0.00, 0.18]	20.4	*
Post-sentence		Agent	not modeled due to insufficient data			
	PROG < PERF	Recipient	PROG: 0.06 [0.00, 0.29] PERF: 0.07 [0.03, 0.15]	0.02 [0.00, 0.05]		
	PROG > PERF	Agent mask	PROG: 0.16 [0.05, 0.34] PERF: 0.04 [0.00, 0.28]	-0.12 [-0.25, -0.03]	108.22	*
	PROG < PERF	Recipient mask	PROG: 0.50 [0.26, 0.74] PERF: 0.74 [0.29, 0.95]	0.24 [0.06, 0.41]	90.17	*

Table 1 shows meaningful effects during the recipient IP, in which there were more fixations to the agent (Bayes factor: 29.77) in the progressive condition (compared with the perfect condition) and more fixations to the recipient (Bayes factor: 97.11) in the perfect condition (compared with the progressive condition); during the theme IP, in which there were more fixations to the agent mask (Bayes factor: 50.95) in the progressive condition (compared with the perfect condition) and more fixations to the recipient (Bayes factor: 19.58) and the recipient mask (Bayes factor: 20.4) in the perfect condition (compared with the progressive condition); during the post-sentence region, in which there were more fixations to the agent mask (Bayes factor: 108.22) in the progressive condition (compared with the perfect condition) and more fixations to the recipient mask (Bayes factor: 90.17) in the perfect condition (compared with the progressive condition).

7.6 Discussion

The analysis of the eye-movement data during sentence presentation and in the 2000 ms post-sentence period confirmed the expectations. The experiment yielded two main results: First, participants relied on the grammatical aspect of the verb as a cue when determining the location of the masked object. After hearing a DO construction with *hand* or *pass* marked with the progressive aspect (*Mr. Pink is handing Mr. Green his sock*) (i.e., during the post-sentence region), participants more often fixated on the location next to the agent (the agent's mask) than when they heard a sentence with the perfect aspect, suggesting that they inferred the location of the object next to the agent. This finding can be attributed to a mental representation in which the described transfer-of-possession event is still in progress, since this initial event stage implies that the agent is still active and in possession of the object. In the perfect condition (*Mr. Pink has handed Mr. Green the sock*), participants looked less frequently at the agent's mask and more frequently at the recipient's mask compared to the progressive condition, suggesting that participants were more likely to interpret the transfer of possession in the event as completed (i.e., that the object was transferred from one of the human characters to the other). Altogether, the findings are interpreted as evidence for the retrieval of a mental representation of a

transfer-of-possession event which was temporally modulated by grammatical aspect, and after sentence end, participants retrieved an appropriate state (i.e. location) of the transferred object.

Second, gaze allocation between the visual human characters during the auditory presentation of the recipient reflected that processing of *hand* and *pass* marked with the perfect aspect more frequently elicited anticipatory eye-movements toward the human character in the visual scene corresponding to the recipient of the description (see IP 'Recipient' in Table 1). This finding suggests that processing a verb marked with aspect allows participants to pre-activate states of referents, i.e., the states of a recipient before or after the recipient has obtained possession of a transferred object.

However, one could argue that the greater dispersion of fixations between agents and recipients in the perfect condition compared to the progressive condition could simply reflect that the agent's activity was less salient and that fixations to the recipient were a superficial by-product of lesser fixations to the agent. In this explanation, fixations to the 'non-agent' visual human character did not reflect an understanding of this as a recipient. Instead, simply hearing *is* or *has* could have modulated attention to the agent, regardless of the semantics of the verb. To address this question, the results of Experiment 1 were compared with those of Experiment 2, in which the lexical verbs were removed and replaced by novel verb, letting listeners rely only on aspect morphosyntax. If hearing only *is* or *has* would modulate the meaning of an activity performed by the agent to a greater or lesser extent, we should see similar effects in Experiment 2, in which this morphosyntax is combined with a novel verb.

Here, two compatible explanations are proposed for the general tendency to fixate the recipient mask more strongly than the agent mask: First, the finding might reflect a tendency toward holistic event interpretation, in which the final event stage was conceptualized but constrained by the progressive aspect. Second, a recency bias (cf. Arnold, 1996) could be responsible for a more salient representation of the recipient referent and, indirectly, of the recipient mask, since the recipient referent had been mentioned more recently than the agent referent at the time participants decided on the location of the object (the theme of the auditory sentence).

8 Experiment 2

In Experiment 2, Experiment 1 was replicated, but all conventional verbs from Experiment 1 were replaced with pseudo-novel, denominal DO verbs, such as *Mr. Pink is clothespinning Mr. Green the sock* (progressive condition) or *Mr. Pink has clothespinned Mr. Green the sock* (perfect condition) (for a description of the choice of linguistic material, see 8.2.). The interest was whether aspect markings of pseudo-novel, denominal verbs in DO constructions could elicit similar effects as in Experiment 1. Therefore, the interest was to see whether participants could infer transfer meaning without accessing a lexical representation for the sentence verb, i.e., by relying on the syntax of the sentence alone.

As in Experiment 1, it was expected that aspect would manipulate the understanding of the transfer event as ongoing or completed, which would be reflected in eye-movements to masks next to the agent and recipient.

8.1 Participants

20 native English speakers participated. The participants were students aged 18-35 recruited from the University of Aberdeen and Robert-Gordon University. Prior to the experiment, participants completed a questionnaire stating whether they knew any languages other than English. Participants who had been exposed to languages other than English before the age of 5 were rejected. Additional verbal questions were asked to confirm whether their language background met the linguistic requirements. Each participant received compensation of £4.

8.2 Materials and design

The experiment used the same stimuli and trial structure as Experiment 1, except all experimental sentences were replaced by DO constructions containing denominal verbs

marked with the present progressive or present perfect aspect, e.g., *Mr. Pink is clothespinning Mr. Green the sock* or *Mr. Pink has clothespinned Mr. Green the sock* or *Ms. Blue is butter knifing Ms. Yellow the cheese* or *Ms. Blue has butter knifed Ms. Yellow the cheese*. The verbs of the filler sentences were also replaced by denominal verbs, e.g., *Ms. Green is oven mitting the pot* or *Ms. Green has oven mitted the pot* and *Mr. Yellow is tin canning Mr. Blue* or *Mr. Yellow has tin canned Mr. Blue*.

Since the sentences referred to the same visual characters as in Experiment 1, Experiment 1 and 2 were identical to the point of the verb used in experimental sentences and filler sentences. The non-human referents presented in each trial consisted of two separately described objects whose affordances were chosen in a way that allowed participants to integrate one as a manner of performing an action, e.g., a clothespin could be used to move a sock and a butter knife could be used to move a piece of cheese (see Kaschak & Glenberg, 2000). This additional object (e.g., the clothespin and the butter knife) was used in order to avoid introducing new, potentially distracting linguistic material during the presentation of the experimental sentence. Importantly, the novel, denominal verbs were invented to not resemble any conventional verbs conveying transfer-of-possession.

8.3 Task

The task was the same as in Experiment 1.

8.4 Procedure

The procedure was the same as in Experiment 1.

8.5 Results

8.5.1 Data pre-processing and modeling

Interest areas and interest periods were defined in the same manner as for Experiment 1. The same modeling approach was used as for Experiment 1.

8.5.2 Visual inspection of the data

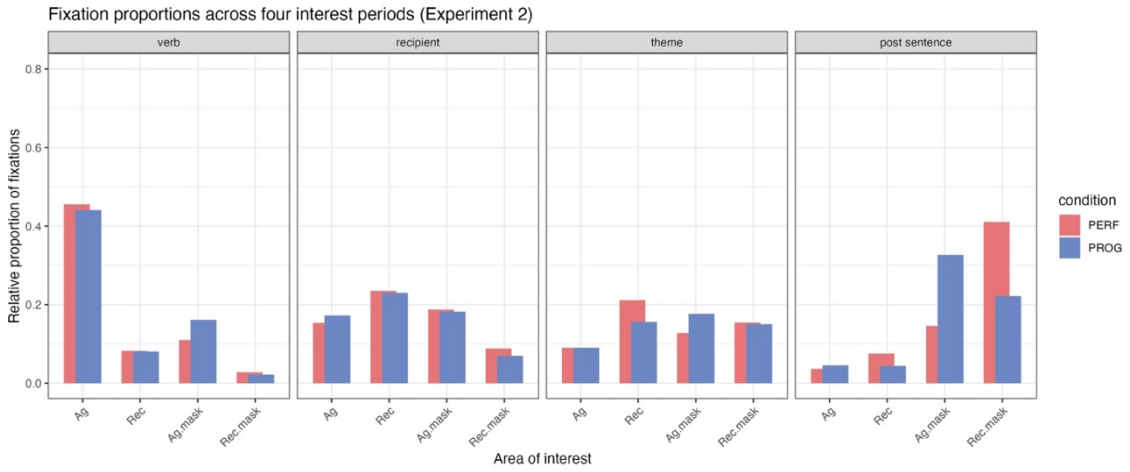

Figure 4. Experiment 2. Relative proportions of fixation in four interest periods (verb, recipient, theme, post-sentence) per critical AOI and condition (aspect); fitted values.

Figure 4 shows the relative fixation proportions in both conditions for all interest areas in the four interest periods. Panels 'verb' and 'recipient' suggest almost no differences between conditions during the first interest periods (verb onset to recipient onset and recipient onset to theme onset). Panel 'theme', which shows the distribution of fixations during the third interest period (theme onset to theme offset), suggests more fixations to the recipient interest area in the perfect condition compared to the progressive condition. Panel 'post sentence', visualizing the distribution of fixations in the post-sentence period, suggests more fixations on the agent mask in the progressive condition and more looks to the recipient mask in the perfect condition.

The same modeling approach was used as for Experiment 1 (Table 2).

Table 2. Experiment 2. A star in the right-most column indicates at least a meaningful difference between conditions, given the data and the model specifications; the higher the Bayes factor the more evidence for the tested hypotheses.

Interest period	Hypothesis tested	Comparison of AOI	Estimated intercepts (mean with 95% CI in [])	Estimated mean difference with 95% CI in []	Bayes factor	
Verb	PROG > PERF	Agent	PROG: 0.44 [0.28, 0.60] PERF: 0.45 [0.22, 0.70]	0.01 [-0.07, 0.09]		
	PROG < PERF	Recipient	PROG: 0.08 [0.04, 0.15] PERF: 0.08 [0.02, 0.25]	0.00 [-0.04, 0.04]		
	PROG > PERF	Agent mask	PROG: 0.16 [0.09, 0.24] PERF: 0.11 [0.03, 0.25]	-0.05 [-0.01, 0.01]	28.62	*
	PROG < PERF	Recipient mask	PROG: 0.02 [0.00, 0.06] PERF: 0.03 [0.00, 0.17]	0.01 [-0.01, 0.03]		
Recipient	PROG > PERF	Agent	PROG: 0.17 [0.10, 0.26] PERF: 0.15 [0.05, 0.34]	-0.02 [-0.08, 0.04]		
	PROG < PERF	Recipient	PROG: 0.23 [0.13, 0.37] PERF: 0.23 [0.09, 0.49]	0.00 [-0.06, 0.07]		
	PROG > PERF	Agent mask	PROG: 0.18 [0.10, 0.29] PERF: 0.19 [0.07, 0.41]	0.01 [-0.06, 0.07]		
	PROG < PERF	Recipient mask	PROG: 0.07 [0.03, 0.12] PERF: 0.09 [0.03, 0.24]	0.02 [-0.01, 0.05]		
Theme	PROG > PERF	Agent	PROG: 0.09 [0.05, 0.15] PERF: 0.09 [0.02, 0.25]	0.00 [-0.05, 0.05]		
	PROG < PERF	Recipient	PROG: 0.15 [0.07, 0.29] PERF: 0.21 [0.05, 0.50]	0.06 [-0.02, 0.15]		
	PROG > PERF	Agent mask	PROG: 0.17 [0.09, 0.29] PERF: 0.12 [0.03, 0.32]	-0.05 [-0.11, 0.01]		
	PROG < PERF	Recipient mask	PROG: 0.15 [0.06, 0.29] PERF: 0.15 [0.03, 0.48]	0.00 [-0.08, 0.08]		

Post-sentence	PROG > PERF	Agent	PROG: 0.05 [0.24, 0.08] PERF: 0.04 [0.0, 0.10]	-0.01 [-0.03, 0.01]		
	PROG < PERF	Recipient	PROG: 0.04 [0.02, 0.08] PERF: 0.07 [0.02, 0.22]	0.03 [0.01, 0.06]	151.81	*
	PROG > PERF	Agent mask	PROG: 0.32 [0.22, 0.46] PERF: 0.14 [0.05, 0.36]	-0.18 [-0.26, -0.09]	518.48	*
	PROG < PERF	Recipient mask	PROG: 0.22 [0.12, 0.37] PERF: 0.41 [0.15, 0.73]	0.19 [0.07, 0.3]	163.61	*

Table 2 shows meaningful effects during the verb IP, in which there were more fixations to the agent mask (Bayes factor: 28.62) in the progressive condition (compared with the perfect condition). During the recipient and theme IP, there were no meaningful effects. During the post-sentence region, there were more fixations to the agent mask (Bayes factor: 518.48) in the progressive condition (compared with the perfect condition) and more fixations to the recipient (Bayes factor 151.81) recipient mask (Bayes factor: 163.61) in the perfect condition (compared with the progressive condition).

8.5.3 Comparison between the results from Experiment 1 and 2

To further evaluate the contribution of verb semantics to the overall interpretation of the sentences, we compared fixations between Experiment 1 and 2 that were recorded in the agent mask and recipient mask AOIs during the post-sentence IP, regardless of aspect. Data was aggregated by subject. The model was set up in a similar way as in the previous analyses. 'Experiment' was specified as fixed effect; a random effects term was included for 'subject' (formula: n | trials(total) ~ 1 + experiment + (1 | subject), where 'n' is the sum of fixations in the recipient mask AOI aggregated over all trials per subject, and 'total' is the total number of fixations aggregated over all trials per subject; dummy-coding: experiment 1 = 0, experiment 2 = 1). All other specifications were as above. Convergence and stability were also assessed in the same way as above. Table 3 reports differences on the probability scale (applying the brms internal inv_logit_scaled-function on

the posteriors) and the corresponding 95% credible intervals (CI). Given the model specifications and given the data, we consider a difference as meaningful, if the difference is estimated to be different from 0, and 0 does not fall within the 95% CI. In addition, we provide the Bayes factor for the hypotheses that were tested.

Table 3. A star in the right-most column indicates at least a meaningful difference between conditions, given the data and the model specifications; the higher the Bayes factor the more evidence for the tested hypotheses.

Interest period	Hypothesis tested	Comparison of AOI	Estimated intercepts (mean with 95% CI in [])	Estimated mean difference with 95% CI in []	Bayes factor	
Post-sentence	Exp 1 < Exp2	Agent mask	Exp1: 0.11 [0.06, 0.18] Exp2: 0.25 [0.06, 0.59]	-0.14 [-0.25, -0.03]	65.44	*
	Exp 1 > Exp2	Recipient mask	Exp1: 0.64 [0.50, 0.76] Exp2: 0.34 [0.11, 0.66]	0.30 [0.15, 0.45]	941.86	*

Modeling the data in the abovementioned way suggests a difference between fixation proportions in the agent mask and in the recipient mask when comparing between both experiments. Participants in Experiment 1 were more likely to direct their gaze towards the recipient mask and less likely to fixate the agent mask than in Experiment 2. This suggests that there was an overall greater bias towards fixating the recipient mask in Experiment 1 compared to Experiment 2.

8.6 Discussion

Consistent with our expectations, the analysis of the post-sentence interest period shows that participants directed their visual attention more often to the agent mask and less often to the recipient mask after hearing *Ms. Yellow is rubber gloving Ms. Blue the chemicals* (progressive condition) compared to sentences such as *Mr. Blue has dining chaired Mr. Yellow the doll* (perfect condition). The results are interpreted as evidence that participants interpreted transfer-of-possession meanings from DO constructions containing novel, denominal verbs and temporally modulated the event by relying on aspectual markings. The interpretation builds on the reasoning that the specific syntactic

configuration with two human referents and one non-animate referent was mapped onto a thematic structure containing an agent, a recipient, and a theme. Since aspect influenced participants' association of the location of the object with the referent denoted by the first post-verbal object of the auditory sentence, we hypothesize that this referent was understood as a recipient as opposed to, for example, a beneficiary. Thus, the aspectual modulation was crucial to distinguish physical transfer of possession from other interpretations. Moreover, this interpretation could not be derived from the nouns and verbs of the sentence, but through an analysis of the syntactic configuration of referents with certain semantic properties.

A question is whether participants may have conceptualized a simpler representation of any given event involving only an agent or an agent and a patient, whereby the progressive aspect increased attention to the agent's activity, causing this referent's mask to receive more visual attention. At the same time, the increased visual attention to the mask of a potentially conceptualized patient in the perfect condition could have reflected a higher activation due to the fact that this referent was mentioned recently. This alternative explanation would imply that recency biases of the patient would be greatly reduced by an agent's ongoing activity and that fixations on masks did not reflect any understanding of possession, but rather a more superficial 'best guess' among three masks. This is partly supported by the finding that aspect did indeed modulate visual attention to the agent's mask during verbal presentation to some extent.

However, recall a potential alternative explanation discussed in Experiment 1, namely that the aspect morphology *is* could have drawn attention to the activity the agent was performing (compared to *has*). In this explanation, fixations to the non-agentive visual referent during the recipient interest period would simply be a random by-product of fewer fixations to the agent. If this explanation is reflected in our data in Experiment 1, then similar effects would be expected during the sentence presentation in Experiment 2. However, after the presentation of the verbs in Experiment 2, no effect of aspect markings of novel verbs was found. It is hypothesized that this is due to the lack of lexical content in the verb that could be used to infer a sentence meaning, which could be further modulated temporally. The fact that no such effect was found in Experiment 2 suggests that

participants in Experiment 1 were indeed analyzing the semantics of the verb in addition to the aspect morphology, whereby a recipient was more likely to be predicted in the perfect condition compared to the progressive condition.

In Experiment 2, gazes were allocated in a way that reflected the processing of aspectually modulated representations of transfer-of-possession events only after the full argument structure of the sentence was available to the listener. For this reason, it seems unlikely that participants did not include the entire sentence in their interpretation. If the human referred to by the first post-verbal argument had been understood as (e.g.) a patient, hearing a given aspect marking should have led participants to launch predictive eye-movements to this visual referent after the verbal presentation in the same way as in Experiment 1. However, there is no evidence in the fixation proportions suggesting any incremental processing of the novel, denominal verbs. Moreover, it would not explain why a patient's *mask*, which was intended to indicate the position of an object, would be fixated after sentence end.

This still leaves the question open as to why the tendency in Experiment 1 of fixating the recipient mask during the post-sentence region were reduced in Experiment 2. We argue that this finding could stem from a varying comprehensiveness of processing between sentences with conventional and novel verbs. Specifically, it is assumed that the sentences containing novel verbs were processed more superficially because of the lack of verbal semantics. Hence, since the recipient is a defining component of the transfer-of-possession meaning, the overall lower fixation amount to the recipient mask in Experiment 2 could indicate a reduced activation of transfer-of-possession meaning compared to Experiment 1.

9 An overview and discussion of the experimental results

9.1 Summary of the experimental results

The study set out to investigate (i) whether listeners use aspect cues to predict and integrate states of referents associated with different temporal stages of transfer-of-possession events and (ii) whether aspect markings of novel, denominal verbs can be used to discern structural from lexical processing.

In two experiments, native speakers of English listened to DO constructions containing conventional DO verbs, e.g., *Ms. Pink is handing Ms. Green the chemicals*; Experiment 1) and novel, denominal verbs, e.g., *Ms. Pink is rubber gloving Ms. Green the chemicals*; Experiment 2) marked with either the present progressive aspect or the present perfect aspect. Their task was to track the location of the object referred to in these sentences (e.g., the chemicals). This object could be moved from a central location to the masks that were placed next to the two human referents in the display (agents and recipients). In line with expectations, the results from Experiment 1 show that participants were more likely to fixate a masked object location next to the agent (Ms. Pink) in a 2000 ms time window after hearing a sentence such as *Ms. Pink is handing Ms. Green the chemicals*, and more likely to fixate a masked object location next to the recipient (Ms. Green) after hearing a sentence such as *Ms. Pink has handed Ms. Green the chemicals*. This difference in fixations to the human characters was observable as soon as during the presentation of the recipient (i.e., prior to the mentioning of the theme). In Experiment 2, eye-movements during the 2000 ms time window showed that participants were more likely to fixate a masked object location next to the agent (Ms. Pink) in a 2000 ms time window after hearing a sentence such as *Ms. Pink is rubber gloving Ms. Green the chemicals*, and more likely to fixate a masked object location next to the recipient (Ms. Green) after hearing a sentence such as *Ms. Pink has rubber gloved Ms. Green the chemicals*.

9.2 Event representations are dynamic in nature

In the experiments reported above, subtle aspectual distinctions of descriptions of transfer-of-possession events were used to modulate event structure and guide overt visual attention to object locations in a visual scene associated with time-related features of event entities. The data suggests that aspect can trigger inferences about states of referents (minimally reflecting a 'before' and an 'after' a recipient has obtained possession), and that the inference of referents' states can occur prior to the referents being mentioned. Furthermore, the study demonstrates that event schemata can be inferred during structural processing, i.e., without lexical access. By focusing on the internal, temporal structure of events, our findings suggest that not only does verb semantics guide eye-movements to referents that fulfill selectional restrictions of verbs (cf. Altmann & Kamide, 1999) and to masked locations reflecting retrieval of object states based on contextual cues from previous discourse (cf. Altmann & Kamide, 2009), but that verb aspect can guide eye-movements to locations in visual scenes associated with states of objects (i.e., before and after an object has been moved). Moreover, anticipatory eye-movements show that verb aspect can trigger the prediction of referents prior to their mentioning.

In a broader theoretical context, the findings support the idea that the cognitive representation of events is temporally complex and cannot be explained alone by the linguistic conception of events as sets of thematic roles such as agents, recipients, and themes. Instead, it seems more appropriate to view event structure as comprising qualitative properties of referents in relation to certain temporal stages (cf. Altmann & Ekves, 2019; Klein; 1999; 2002). For example, an agent is only necessarily active until a recipient has obtained possession, and a recipient only qualifies as such after having obtained possession. This implies that the qualities and salience of referents differentiate over time in accordance with the currently perceived visual or linguistic input. In this study, we demonstrated that participants were able to track the extrinsic state of an object changing in location between two human referents, suggesting that comprehending DO verbs and DO constructions marked with grammatical aspect involves retrieving object states associated with certain temporal event stages.

A question arises as to whether our findings reflect the mental representation of an object changing state over time (cf. Altmann & Ekves, 2019), or whether the findings simply can be attributed to a more superficial allocation of mental attention between agents and recipients, by which aspect solely influenced the salience of referents. The latter explanation would imply that aspect increased salience of agents and recipients without regards to temporal features of events, allowing participants to subsequently logically infer the object location, as opposed to retrieving a particular object state in online processing. It can be argued, however, that both explanations (salience of referents vs. object state retrieval) may only appear to differ at first glance. If we look exactly at what would cause increased salience of a referent when a certain aspect marker had been heard, it becomes apparent that it is the quality of being in possession of an object at a specific event time which allows the perceptual system to discern two possible possessor candidates. In other words, identifying a human referent as a possessor is correlated with identifying a specific state (i.e., location) of a transferred object (the theme), meaning that the inference of the state of the transferred object must have been a prerequisite of identifying the possessor.

This interpretation goes in line with the argumentation laid out by Altmann and Ekves (2019). According to their theory of event representation, perceiving an event through language involves, in the simplest case, activation of multiple competing states of the same object. An a priori argument for multiple-state activation is that the changes defining events are only understood as such by discerning qualities of an object associated with a 'before' and 'after' the object changed. Simply recognizing a single state of an object would not suffice to understand that a change has occurred or will happen; this understanding presupposes the simultaneous competing activation of a previous or subsequent state that will eventually be suppressed in final interpretation. The data from Experiment 1 and 2 gives us indications about how multiple object states, i.e., mentally represented states of schematic event entities are retrieved in association with different temporal event stages.

9.3 Processing of sentences containing novel verbs

Listeners in Experiment 2 did not generate any expectations about upcoming referents during or shortly after presentation of the novel verbs, presumably since these verbs did not provide access to lexical information containing syntactic specifications. Only after sentence end, eye-movements suggest that listeners were able to derive and temporally modulate representations of transfer-of-possession events by analyzing sentence forms and aspectual markings in the absence of lexical verbs. However, whereas it cannot be ruled out that post-hoc processing involved retrieving a lexical item (e.g., *give*) that would guide interpretation, this process would still rely on an analysis of syntactic configurations.

One could object that if participants were relying on structural information alone to derive sentence meaning, we would expect eye-movements revealing sentence interpretation during the unfolding of the sentence containing a novel verb. However, if we look at models of interpretation of novel verbs (specifically denominal verbs; see Michealis & Hsiao, 2021; *co-composition*, Pustejovsky, 1995; *conceptual blending theory*, Fauconnier & Turner, 2004), they share the assumption that the meaning of novel verbs is accommodated by the semantic restrictions associated with the arguments structures in which they appear. From an incremental point of view, this means that processing the first post-verbal argument would induce a causative interpretation, followed by modifying this interpretation to caused-possession upon integration of the final argument conveying a referent that fulfilled the selectional restrictions of transfer-of-possession schema. Returning to the before-mentioned issue, these models would predict that if participants were relying solely on structural processing, participants would derive a causative event schema after hearing a novel verb and before hearing the second post-verbal argument. Whereas our eye-movement data suggests that participants only landed at an interpretation after sentence end, it is important to keep in mind that if participants would have interpreted a causative event from hearing a novel verb and the first post-verbal argument, this would not necessarily be reflected in eye-movements to referents or masks. Since the ongoing activity of an agent and the state of a patient in causative event such as ‘Laura pinched

her sister' do not involve state changes that occupy different spatial locations over time (i.e., across their before- and after-states, as in the transferring of an object), the current experiment design did not provide participants with spatial regions to fixate that would correlate with before- and after-states of referents of a causative events. Hence, our data leaves the question open as to whether participants did indeed incrementally interpret sentences as they were unfolding and relied solely on structural information. However, the data suggest that structural processing as a minimum played a role in the inference of sentence meaning, regardless of the involvement of the retrieval of a lexical item during post-hoc processing. However, the exact involvement of lexical processes is a concern for further studies investigating online language comprehension. Moreover, since the current study investigates how listeners are able to derive sentence meaning from syntactic configurations in comprehension, it provides limited insight in whether abstract sentence meanings are available to speakers independently from lexical access (see Wheeldon & Konopka (2023) for a discussion on the role of lexical and structural processes in the context of production and comprehension).

Interestingly, the comparison of results from Experiment 1 and 2 shows that there was an overall greater recipient bias (i.e., higher fixation amount to the visual referent corresponding to the recipient of the auditory sentence) when a double-object sentence containing a conventional verb was heard compared to a novel verb. We propose an explanation for these findings that goes in line with basic principles found by Pickering and Branigan (1998) and Cleland and Pickering (2003), who demonstrated that syntactic specifications licensed by verbs can superimpose abstract, generalized sentence structures and enhance re-activation of these structures in subsequent language production. Since our results show that abstract structures are analyzable in the absence of lexical access to verbs, processing a conventional verb in addition to an abstract sentence form should activate overlapping syntactic (and thereby semantic) representations and cause mutually reinforced activation. In this explanation, the comprehension of the meaning of double-object sentences should receive an increased activation when containing double-object verbs. It is possible that this overlap caused a boost in event schema activation, thereby

accounting for the recipient bias evident in Experiment 1, since conceptualizing a recipient is a defining component of deriving a transfer-of-possession event schema.

9.4 Comparison with earlier methods

In line with earlier findings showing recipient biases in pronoun resolution (Arnold, 1996; Stevenson, Crawley, & Kleinman, 1994), the results of the present study show a tendency of understanding recipients as possessors at the final stage of transfer-of-possession events. However, in Rhode, Kehler and Elman's (2006) study on the resolution of ambiguous pronouns following prepositional ditransitive sentences marked with simple past tense or the past progressive aspect (*John handed/was handing a book to Bob*), participants tended to maintain a bias of re-mentioning agents (*John*) across conditions. The differences from our results may stem from the use of the present perfect aspect (e.g., *has handed*) in our experimental stimuli, which could have increased salience of the recipient by highlighting a final event stage. On the other hand, the simple past tense (e.g., *handed*) might have elicited a holistic interpretation without accessing internal temporal properties of the described events, leaving the salience of internal event components (agents and recipients) less definitely specified and thereby allowing the readers to interpret more freely on the internal properties of the event structure.

The findings of aspect sensitivity in the interpretation of ambiguous pronouns led Rhode, Kehler and Elman (2006) to conclude that pronoun resolution not only draws on thematic configurations, but that it is additionally sensitive to temporality on an event structural level. However, the authors did not address the issue that pronoun interpretation biases could stem from a more superficial statistical correlation of the frequency of aspect use in prepositional ditransitive sentences and experienced anaphoric resolutions in subsequent discourse. Hence, the effects could be attributed to the reader's implicit probabilistic processing of the likelihood of a pronoun referring to either the agent or the recipient when presented with a specific aspectual marking. Considering this alternative explanation of their data, processing the grammatical aspect markings and the configuration of thematic roles would indeed suffice to solve the task, as opposed to drawing on temporally

modulated event representations. In our study, however, we did not focus on effects of salience of event entities, but instead let participants track the location of the transferred object. Since participants had to map comprehended language onto a visual scene, participants could not rely on the probabilistic of the co-occurrence of linguistic means, but instead had to retrieve mental representations serving to mediate between linguistic and visual perception. Furthermore, by considering prediction effects, the results of the current study cannot be explained by associations on a superficial linguistic level, but instead indicate expectations of the integration of visual referents into a temporally modulated event representation.

The same issue arises from the method used by Grüter et al. (2018) that revealed interdependence between aspect markings in DO constructions and co-reference processing. Based on these findings, the authors argued that listeners' anticipatory fixations to agents and recipients prior to their anaphoric mentioning reflected access to temporally modulated representations of transfer-of-possession. However, this method still leaves open the possibility that a more superficial analysis of thematic roles and aspect could trigger an implicit probabilistic processing of the most likely discourse continuation. Again, this would not necessarily reflect access to temporally modulated event representations, but solely the choice of a visual referent that fulfilled a semantic criterion (in this case, the gender of a referent) of a thematic role most frequently experienced in discourse after a certain aspect use.

A further issue of their study arises from only measuring the anticipatory fixations in relation to the salience of human referents. This allows for a processing strategy which involves simply attributing the agentive entity with an activity of varying salience depending on the aspect used. By considering this potential confound, it cannot be ruled out that fixations to the recipient (which in this explanation simply would be an alternative referent to the agent) could be the by-product of less fixations to the agent, regardless of any in- or decreases in salience of the recipient. The plausibility of participants having developed this processing strategy is further increased by the linguistic material used in the study. The verbs used denoted a wider range of transfer-of-possession events by which guaranteed transfer is either not implied (*roll* and *throw*) or does not involve physical

possession (as is the case in beneficial meaning denoted by present) (cf. Rhode, Kehler & Elman, 2006; cf. Beavers & Koontz-Garboden, 2020). However, as previously pointed out, in the current study, not the relative salience of agents and recipients was investigated, but instead, participants tracked the location of a transferred object.

10 General discussion

10.1 Summary of the theoretical and empirical study

The main goal of this dissertation was to investigate the representation and processing of events, focusing on their subtype involving human actions and the temporal and multi-layered nature of these. The investigation examined the hierarchical and temporal structure of events, using cascade theory, Argument-time Structure Theory (ATS), and Intersecting Object Histories (IOH) to model different aspects of cognitive representation. These frameworks were chosen for their complementary approaches to capturing logical and temporal structures of events. In addition, the linguistic encoding of events was examined through constructional approaches to language and the use of grammatical aspect. The former aimed to understand how abstract syntactic constructions, such as the double-object construction, encode event meaning independently of specific lexical items. The latter focused on understanding how the temporal perspectivization of an event, conveyed through grammatical aspect, is cognitively processed. ATS offers a framework for exploring how such temporal structures are linguistically accessed and represented, particularly through aspectual marking in the English language.

An essential insight of the study is that actions and events are multilayered, as the tokenization of a human action implies the simultaneous tokenization of multiple interrelated actions, each forming a distinct layer in cognitive representation. For example, when one is performing the act of handing money to a cashier, one is simultaneously performing the act of paying for a good (given that certain conventional conditions are met). Both actions coincide in time and space. Our understanding of paying for an item thus implies the understanding of handing over money, and both actions form layers of the same event. More fine-grained, this implies further layers, for instance the layers involving the act of moving an arm to carry money and the layer involving having the intention to do so.

Closely related to these assumptions are the approaches laid out by IOH and ATS, which capture the dynamic nature of actions and events in terms of their logical and

temporal complexity (Altmann & Ekves, 2019; Klein, 1994; 1999; 2002; 2010). According to these theories, events are defined by the changes that their involved entities undergo. In the case of events involving human action, these changes are caused by the action itself. Recognizing state change as the defining criterion for events (as opposed to static states without change) implies that events unfold over time, comprising multiple object states interconnected by counterfactual and causal dependencies. For example, the transfer of an object from one person to another unfolds over time, involving a 'before' and an 'after' where the recipient gains possession of the object. It is implicitly understood that the recipient did not possess the object prior to the transfer. This counterfactual dependency is central to the cognitive representation of the event, as co-representing the prior object state is essential for recognizing that a state change has occurred.

Regarding how events are linguistically encoded, there are a priori arguments suggesting that a syntactic structure, such as the double-object construction, can associate meaning independently of the verb that instantiates it (cf. Goldberg, 1995). In addition, Klein's (2002) analysis shows how the internal temporal structures of an event can be accessed through the finiteness in the English language.

To empirically test the theoretical insights concerning the linguistic encoding of events, two experiments were conducted to examine how abstract syntactic structures and grammatical aspect influence real-time event comprehension. Using the visual-world paradigm, Experiment 1 investigated whether aspect markers (e.g., present progressive vs. present perfect) guide listeners' eye movements toward locations in visual scenes corresponding to different temporal phases of transfer events. Sentences with ditransitive verbs (e.g., *hand*, *pass*) were used to determine whether listeners predict and integrate information about the states of referents during sentence processing. The results showed that grammatical aspect modulates eye movements toward givers and receivers based on whether the transfer is perceived as ongoing or complete, as conveyed by the progressive and perfect aspect. Notably, eye movements reflected aspect-driven predictions of referents before they were explicitly mentioned.

Experiment 2 extended these findings by exploring whether the double-object construction, even when it does not contain a conventional ditransitive verb, activates a

transfer event schema that can be modulated by grammatical aspect. Pseudo-novel verbs marked with progressive or perfect aspect were used (e.g., "Mr. Pink is *rubber glove*-ing/has *clothespin*-ed Mr. Green the sock") to test whether listeners infer transfer-of-possession meaning from the syntactic structure alone. The results demonstrated that listeners rely on the double-object syntax to construct transfer event representations, and that grammatical aspect dynamically shapes these representations, eliciting similar effects to those observed in Experiment 1. These findings highlight that abstract syntactic structures and aspectual morphosyntax interact to activate and temporally modulate meaning representations during real-time language comprehension.

10.2 The empirical findings in the context of cascade theory

A central insight from the empirical study was how temporal properties of events are processed in real-time event comprehension. However, cascade theory offers a complementary perspective for interpreting the findings. By emphasizing the multilayered nature of human actions and events, it reveals how dynamic temporal structures, particularly event phases, are embedded within hierarchical layers of cognitive representation. This approach integrates causal interdependencies in event representation and explains how linguistic cues, such as grammatical aspect, interact with these event layers. As outlined in chapter 2, cascade theory views events as hierarchical structures composed of interconnected layers, each representing a distinct conceptualization of the event. These layers range from higher causal levels of an action (e.g., paying for an item) to more fine-grained, lower-level actions (e.g., handing over money).

This framework is particularly useful for understanding the experimental results, which demonstrated how grammatical aspect modulates attention to specific temporal phases of events. In the experiments, listeners' eye movements reflected sensitivity to the progressive and perfect aspects of double-object constructions. For instance, the progressive aspect (*is handing*) directed attention to the ongoing transfer, focusing on the dynamic act of moving the object, while the perfect aspect (*has handed*) shifted focus to the resultant state, emphasizing the completed transfer where the recipient possesses the

object. From the perspective of cascade theory, these temporal phases correspond to distinct layers of the event structure. The progressive aspect (*is handing*) highlights a lower layer of the cascade, representing the physical act of transferring the object to the recipient. In contrast, the perfect aspect (*has handed*) highlights a higher layer of the cascade, representing the successful completion of the transfer and the causation of a resultant state. This aligns with cascade theory's emphasis on causal dependencies, where higher layers presuppose and build upon the processes occurring in lower layers. For example, the act of causing someone to receive an object (a higher layer) – depending of the type of giving – cannot occur without the act of moving the object to the recipient (a lower layer). Grammatical aspect, therefore, serves as a linguistic cue that allows listeners to dynamically navigate these event layers during real-time comprehension. By providing cues to focus on specific temporal phases, aspectual markers enable listeners to parse and conceptualize the event cascade in a manner that reflects its hierarchical organization. This interaction between grammatical aspect and event layers underscores the utility of cascade theory for explaining how linguistic structures interface with cognitive representations of events, particularly in contexts where temporal and causal dependencies play a crucial role.

Cascade theory further illuminates the role of counterfactual and causal dependencies in event comprehension, in a manner consistent with Argument-time Structure Theory and Intersecting Object Histories. For an event to be understood as a cohesive whole, listeners must co-represent prior states (e.g., the object's initial position) and future possibilities (e.g., the recipient possessing the object). The experiments showed that listeners anticipate and integrate these dependencies in real time, as evidenced by their eye movements across visual areas associated with different temporal phases. For example, the act-token 'causing someone to receive an object' causally depends on the act-token 'moving the object to the recipient'. This dependency is not only temporal but also hierarchical – understanding the higher-layer action (receiving) presupposes understanding the lower-layer action (moving). In sum, the findings thus highlight the interplay between temporality and multilayeredness, suggesting that event phases are not only sequenced over time but are hierarchically organized as causally interdependent layers.

Cascade theory raises avenues for future research, particularly in understanding the broader implications of hierarchical and multilayered event structures. For example, further investigation is needed into how other linguistic markers, such as tense or modality, interact with event layers. Modality, for instance, plays a role in introducing abstraction and contingency. An expression like *She might hand him the book* situates the event in a hypothetical domain, adding a layer of abstraction detached from any actualized actions or resultant states. Similarly, an expression like *She could hand him the book if he asks* encodes a contingent relationship where the event's realization depends on another action. These examples illustrate how modality not only interacts with the temporal and counterfactual dependencies central to cascade theory but also embeds layers of potentiality and conditions. Additionally, questions remain about the generalizability of these findings to events with more complex causal dependencies or non-linear temporal structures, which could challenge or extend the current cascade framework. As will be described in 10.4, cross-linguistic comparisons also offer valuable insights, as languages with different aspectual systems or syntactic rules may encode and conceptualize multilayered events in ways that reveal further dimensions of cognitive and linguistic interaction.

10.3 Limitations of the experimental study

In the experimental study, it was among other things demonstrated that listeners can rely on syntactic structure to infer sentence meaning. Specifically, the experimental results suggest that listeners derive a transfer understanding of double-object sentences containing novel verbs, i.e., verbs that in themselves cannot contribute with transfer meaning. The study set out to test principles put forward by Construction Grammar, which involve the idea that syntactic structures and semantic structures form symbolic units of form and meaning. This involves the prediction that when the speaker produces an utterance, he or she retrieves a semantic structure that is mapped onto the associated syntactic structure (by which structural processing and functional assignment are conflated, contra the two-stage model of speech production (cf. Dell, 1995; Garret, 1975; see, e.g., Bencini, 2013).

Since the current experimental study investigated mechanisms in language comprehension, it does not provide any details about the mechanisms involved in syntactically encoding semantic structures, or, generally speaking, how the composition of a sentence takes place in the speaker's mind (see chapter 4.2.3). Instead, it takes the listener's perspective and shows what information they can rely on to derive meaning from a linguistic input. This opens up the possibility that it is logical inference, and not mapping between syntactic and semantic structures as such, that forms the basis for understanding. For example, it cannot be ruled out that after hearing a three-participant structure, the listener logically inferred during post-hoc processing that this structure usually occurs with the verb *give*, and that *give* has something to do with transfer.

However, regardless of whether any lexical words associating transfer were retrieved in post-hoc processing, this still had to be done on the basis of the analysis of syntactic structure. Thus, the analysis of the meaning of the sentence must have originated from the analysis of the structure of the sentence, either as a logical inference or as a mapping process of syntax and semantics. Note that this unclarity persists in other studies that have also investigated the processing of structural information from a listener-based perspective (Bencini & Goldberg, 2000; Kaschak & Glenberg, 2000). It is therefore a task for future studies to separate structural from lexical processes in language production (for further reading, see Baumann, Pappert, & Fitz, 2015; Baumann, Pappert & Pechmann, 2021; Chang, Dell, & Bock, 2006; Ziegler et al., 2018).

10.4 Cross-linguistic variation and cascades

As we saw in the theoretical part of the dissertation, cascades model multilayered structures in language and cognition (cf. Löbner, 2021). This concerns on the one hand our multilayered understanding of events involving human actions, and on the other hand the multilayered representation of the semantics of lexical of verbs and syntactic constructions (see chapter 4.3). The basic idea is that an expression generates a particular level in a cascade, allowing subordinate levels to be considered as part of a multilayered representation. On a morphosyntactic level, verb aspect can be considered to generate a level

in the cascade by either highlighting a higher-level resultant state or a lower-level ongoing phase of the event expressed by the verb (see chapter 4.4.2). This means that the English language has morphosyntactic means to assert levels in a cascade that languages such as German do not (as the German language does not provide the option of marking the verb with progressive aspect).

The idea that different languages per linguistic conventions can assert different levels of a cascade has already been put forward by Löbner (2020; 2021). Among Löbner's (2021: p. 286) examples of this are the German verb particles *tot-* ('dead') and *er-* (a resultative particle), which can be attached to verbs expressing various ways of killing, e.g., *erschießen* or *totschießen* (both with the meaning 'shoot to death') (see Löbner's and Van Valin and LaPolla's (1997) analysis of the verb *kill* in chapter 4.3.1). These verb particles show that German has morphological means (e.g., *er-*) to achieve a level of result in a cascade that English does not.

If we look at how English, French, and German speakers describe motion events, similar cross-linguistic variations are evident. Stutterheim, Lambert, and Gerwien (2021) found that while English speakers primarily and German speakers almost exclusively use manner verbs to describe motion events (e.g., *gehen* 'walk', *laufen* 'run'), French speakers only use manner verbs in about half of expressions, with the rest being path verbs (see chapter 10.3.1.2 for a more detailed discussion on these variations). If we recall the cascade theoretical analysis in chapter 2.3.1, we saw how a motion event can be divided into cascade levels with counterfactual relationships between an event layer 'motion' at the lowest level, which generates a higher level 'manner of motion', which further generates a higher level 'path of motion'. The rationale for the hierarchical classification is that one can only move in a certain way by moving at all, and one can only move towards a goal by moving in a certain way (see also Talmy's (2000) definition of manner as a co-event of motion). This shows that since languages by convention vary in whether they include path or manner in an assertion, it can be assumed that speakers of different languages assert different levels in a cascade. The cascade-theoretic modeling of multilayered structures thus offers a clearer insight into what falls under a representation of, e.g., a motion event in the minds of speakers across languages.

10.5 Challenges for Construction Grammar

Construction Grammar played a central role in this study by positing the association between syntactic structures and event structures, an idea discussed in the theoretical chapter and tested in the empirical chapter. It provided a perspective instrumental in exploring how syntactic constructions encode event meaning independently of specific lexical items, laying a foundation for testing their role in event comprehension. However, as discussed in chapter 4, basic tenets of the theory have faced critiques regarding its consistency in applying to all linguistic expressions. Additionally, it has been criticized in the domain of language acquisition, particularly for relying on inductive learning and on the entrenchment of linguistic patterns through input frequency. While these challenges lie outside the direct focus of this study, they will now be considered to provide a more comprehensive view of the framework. Additionally, they will help suggest future directions for validating its robustness and appropriateness as a model of linguistic representation.

A critique of Construction Grammar – and of usage-based theories more broadly – centers on the explanation of how language is learned. In Construction Grammar (e.g., Goldberg, 2006; 2019), it is assumed that language acquisition is based on inductive learning. Each experience with a linguistic pattern activates a node (or a network of nodes), which through high frequency of input can become entrenched and facilitate further activation (see Tomasello, 2003). Entrenchment – along with cognitive mechanisms such as categorization and analogy (see Diessel, 2015) – is assumed to be based on general cognition, rendering any idea of an innate linguistic faculty redundant (see, e.g., Chomsky, 1965). Instead, it is the bottom-up experience with language which is forms the basis for the recognition of recurring patterns which ultimately can lead to grammatical representation.

Hence, the idea of frequency-based, inductive learning is central to a Construction Grammar approach to language acquisition. In addition, it is assumed that the frequency of the input plays a role in a parallel manner both when learning the first and the second language (Wulff & Ellis, 2018). However, according to recent studies in second-language

acquisition, there is an asymmetry between how the first and second languages are learned (Goldberg, 2019; Stutterheim, Lambert & Gerwien, 2021). This asymmetry stems from transfer effects from first-language grammar and lexicon which cause the production of unidiomatic or ungrammatical expressions in the second language, even though the correct expressions in the target language can be assumed to be experienced with high frequency. This finding thus poses a challenge to usage-based approaches to language acquisition, and it is necessary for the robustness of Construction Grammar to account for this asymmetry.

In the following chapters, we will look at some principles laid out by Goldberg (2019) that set out to explain this asymmetry in order to preserve a purely usage-based approach. Goldberg introduces cognitive mechanisms (*coverage* and *statistical preemption*) that account for asymmetries between first- and second-language acquisition within a framework based on inductive learning, i.e., learning primarily or solely based on frequency of an input. We will then look at a competing theory laid out by Stutterheim, Lambert, and Gerwien (2021) that explains the same asymmetry based on principles that go against theoretical frameworks that solely rely on frequency of an input. According to the latter study, entrenchment of conceptualization routines forms constraints on how much of a role the frequency of an input plays. Although it will be argued that the two approaches do not necessarily form a strict opposition, the criticism of the usage-based approach poses a withstanding challenge for Construction Grammar that should be addressed in future studies.

10.5.1 The role of statistical learning in language acquisition

As we have seen in chapter 4.2, fundamental to Construction Grammar is a semasiological and usage-based approach to language. It is assumed that language acquisition involves identifying abstract meanings that are shared across instances of syntactic forms. In this process, the frequency of experiencing a form is a driving factor in achieving full competence when both a child and an adult learns a language (see, e.g., Ellis, 2002). However, when comparing how children and adults learn languages, fundamental

asymmetries emerge. A central asymmetry is that the production of erroneous (i.e., ungrammatical or unidiomatic) expressions in the L2 is fundamentally characterized by features of the L1. This transfer between languages goes beyond mispronunciation of the L2 words, but also manifests itself in a choice of linguistic means at a lexical and at a more abstract morphosyntactic level. Even very advanced learners' errors in L2 production bear the imprint of grammatical and lexical features of their L1 in spite of a high-frequency input of the idiomatic or grammatical structures of the target language. This phenomenon thus poses a challenge for a purely frequency-based, i.e., statistical, approach to language acquisition.

In this chapter, we will look at two fundamental perspectives on the role of statistical learning in L2 acquisition. One comes from the Construction Grammar framework and concerns the concepts of coverage and statistical preemption (Goldberg, 2019). These concepts describe mechanisms that underlie the productivity of an expression (i.e., the likelihood of novel usage of an argument structure construction) and the idiomatic use of argument structure constructions in L2 production. Another perspective challenging the role of statistical learning, and hence, Goldberg's assumptions, comes from Stutterheim, Lambert, and Gerwien (2021) who attribute the main factor of erroneous L2 production to entrenched L1 conceptualization routines.

The aim of the chapter is, based on these two different approaches to language acquisition, to identify distinct mechanisms involved in language acquisition (more specifically, in language production in L2), and to shed light on some challenges of a usage-based (i.e., Construction Grammar) approach to language and cognition.

10.5.1.1 Coverage and statistical preemption

We first turn our attention to Goldberg's notion of *coverage* (see chapter 4 in Goldberg (2019)). Coverage refers to the size of the conceptual domain that a construction covers, that is, the degree of flexibility that its meaning has been experienced with. According to this idea, each new verb experienced in an argument structure construction expands the conceptual domain covered by the construction. For example, while the DO construction

can be assumed to frequently be instantiated with *give*, other more or less frequent variants (*send*, *hand*, *pass*) can extend the conceptual domain covered by the DO construction. Hence, the conceptual domain of a construction refers to the experienced number of individual types of a construction (in particular, the number of instantiations with different verbs). The larger the conceptual domain, the more likely the speaker is to productively coin a new expression, i.e., to use a not previously attested verb in the construction. Hence, creative use of a construction is mainly attributed to the degree of flexibility experienced with the construction.

In addition to the number of different types of a construction, the variability (i.e., the similarity) of these types influences how likely it is that the construction will be used in a novel manner. For instance, having experienced the less prototypical verb *bake* in the DO construction will count as evidence for the speaker that this construction allows for a more variable use, thereby extending the domain of this construction (i.e., the flexibility of the use of the construction) and increasing the likelihood of this construction being used with an additional not previously attested verb. At the same time, a greater density in the clustering of variants of a construction (i.e., a high token input of a semantically similar group of types, e.g., *hand* and *pass*) can constrain the productivity of a construction. This means that if a construction has only been experienced with a high similarity of semantic variations, it will less likely allow for new expressions deviating from these narrow semantic variations (despite a potentially high type frequency). In other words, a more evenly distributed coverage of types facilitates the productivity of a construction.

A second important principle laid out by Goldberg is the notion of *statistical preemption* (see chapter 5, ibid., 2019; see also Stefanowitsch's (2008) notion of *negative entrenchment*). Statistical preemption describes how, in language production, an implicit, statistical mechanism monitors for negative evidence, influencing how likely one linguistic structure is to be used over other seemingly grammatically correct but idiomatically misplaced structures. Goldberg's main example in this context is the odd usage of *explain* in the DO construction, as in *?Explain me this*. For Goldberg, this serves as an example of a misuse of the DO construction that children in their L1 acquisition can easily overcome, but adults in their L2 acquisition of the English language show more difficulty

identifying as improper use. Importantly, there is no apparent semantic restriction on why *explain* cannot be used in this construction while, e.g., *tell* can; in contrary, the restrictions in argument realization with this verb are assumed to be a mere arbitrary convention.

Goldberg attributes this observation to fundamental differences between how the L1 and the L2 are acquired (see chapter 6 in ibid., 2019). At the core of these differences is how the conceptual domain of a construction, in this case the DO construction, is shaped by the construction types experienced in L1 acquisition. Since children presumably only experience examples of the idiomatic use of *explain* in the PO construction (e.g., *Explain this to me*), they are able to infer from the indirect negative evidence (hence, that they do not hear *explain* in the DO construction) that the DO construction is not proper use with this verb.

However, according to Goldberg, the keeping track of negative evidence works less effectively in L2 acquisition than in L1 acquisition. When learning the DO construction in L2 acquisition, a transfer from its usage in the learner's L1 can take place, causing interference across languages. As a result, the implicit statistical mechanism monitoring negative evidence (and thereby statistically preempting the unidiomatic use) will be affected by the difficulty discerning which of the two languages the word *explain* cannot be used in the DO construction. This idea can be made clear by transfer effects we can assume to take place between L1 German and L2 English. Since in German, the DO realization of *explain* is permitted (as in *Erklären Sie mir das*), this L1 usage can cause interference for German L2 learners of English.

The concepts of coverage and statistical preemption can thus explain why an L2 learner can have a high token input of a certain type while still producing erroneous expressions. These concepts then allow Goldberg to retain a usage-based framework that involves a strictly statistical (i.e., frequency-based) approach to how language is acquired, by which phenomena that seemingly contradict a purely statistical approach can be explained.

10.5.1.2 Entrenchment of first-language conceptualization patterns

As we have seen in the previous chapter, Goldberg (2019) explains asymmetries between L1 and L2 acquisition by concepts that preserve a strictly statistical approach to language learning. Goldberg's idea of statistical preemption is that negative evidence is harder for an implicit probabilistic mechanism to keep track of as concepts shaped by the L1 can compete and infer with the mechanism, causing the production of unidiomatic (or even ungrammatical) expressions in the L2. By assuming these principles, Goldberg preserves a purely statistical approach to language learning. This approach involves the prediction that, on the one hand, the primary factor of acquiring a language is the accumulation of experiences (the number of tokens experienced), and on the other hand, that learners apply the same statistical learning mechanisms in L2 as in L1 acquisition (whereby asymmetries in L1 and L2 acquisition are explained by inference in these mechanisms) (see Wulff & Ellis, 2018).

However, in a recent study by Stutterheim, Lambert, and Gerwien (2021), asymmetries in L1 and L2 acquisition were investigated with a different approach. Instead of assuming mechanisms explaining apparent violations of statistical learning (thus preserving a frequency-based approach; cf. Goldberg, 2019), this study linked asymmetries in acquisition to the idea that L2 production is influenced by event conceptualization patterns entrenched in L1 acquisition. Specifically, certain ways of construing events are assumed to be entrenched in accordance with syntactic and lexical means of encoding, causing language-specific activation routines of the concepts underlying linguistic form. Entrenchment is thus assumed to provide a rigid framework for language acquisition in adult learners. Hence, in this view, statistical learning is not considered as the main factor of L2 acquisition since entrenched event construal is seen as a constraint on the boundaries within which frequency of an input can play a role.

This idea was experimentally investigated by Stutterheim, Lambert, and Gerwien (2021) by examining how French L1 speakers describe motion events in their L2 German and English. First, the study showed that while German and English L1 speakers primarily use manner verbs to describe motion events (for German, this applies to almost all

descriptions, e.g., *gehen* ‘walk’, *laufen* ‘run’), French L1 speakers use manner verbs in only half of description, while path verbs are attested in the rest of the French L1 descriptions. In addition, while L1 French speakers typically combine manner verbs with locative adverbials (*une femme marche dans la rue* ‘a woman walks in the street’), L1 German and English speakers typically combine manner verbs with directional adverbials (*eine Straße entlang* ‘along a street’; *zu einer Kirche* ‘towards a church’). Based on these L1 elicitations, the authors argued that motion events are construed in different ways across languages: In French, motion events are typically construed within a locative event frame and in German and English within a directional event frame (i.e., the languages differ as to whether a locative or a directional event layer is typically conceptualized; see chapter 3.1).

In a further experiment, the study showed that when French L1 speakers describe the same motion events in their very advanced level of L2 German and English, they maintain their tendency to use manner verbs in the locative event frame. While this usage is not ungrammatical, it clearly differs from the usage of L1 German and English speakers, who typically implement manner verbs in a directional event frame. The authors interpreted this tendency as evidence that entrenched L1 conceptualization patterns are automatically activated when the speaker mentally prepares descriptions of motion events, even when the L2 is used. This language-specific construal forms the basis for the selection of syntactic and lexical means across languages, which causes unidiomatic language use in the L2. Importantly, this happens despite an assumed high token input of the typical language use. Based on this, the authors conclude that it is not only the statistical accumulation of experienced tokens that plays a role in L2 acquisition, but instead that entrenched conceptualization patterns may be a driving factor in acquisition by constraining the effect of input frequency.

10.5.1.3 Language-specificity in conceptual cognitive structures

In the two previous chapters, we have seen two different approaches to explaining asymmetries in L1 and L2 acquisition. One is purely statistical (cf. Goldberg, 2019), while

another competing theory identifies limitations of a statistical approach imposed by entrenched conceptualization patterns in language production (cf. Stutterheim, Lambert & Gerwien, 2021). The latter study demonstrates that when French L1 speakers produce descriptions of motion events in L2 German or L2 English, they conceptualize events according to the means of expression in their native language, resulting in the selection of syntactic and lexical structures that are unidiomatic in the target language. This suggests that the linguistic encoding of events in a particular language influences general conceptualization routines involved in event cognition. Hence, the study counts as evidence against the idea that cognitive representations are universal and consist of linguistically autonomous conceptual structures that can be mapped onto linguistic structures.

Goldberg claims in an apparently similar vein as Stutterheim, Lambert, and Gerwien (2021) that "well-practiced first language warps aspects of the […] conceptual space that includes the representations required for speaking." (2019, p. 140). However, according to Goldberg, the "warping" of the network of conceptual structures native to the L1 is not rooted in entrenched conceptualization routines (i.e., in the automatized retrieval of event frames), but rather in a warped coverage (see chapter 10.3.1.1) of a certain construction, on the basis of which unidiomatic usages of this construction are preempted in L2 production.

This means that when the speaker prepares a conceptual representation (corresponding to an event frame or the semantic side of a construction), Stutterheim et al. assume that the speaker per routine activates an entrenched conceptual structure which leads to activation of the unidiomatic linguistic form. Goldberg, on the other hand, assumes that the speaker does not per routine activate an event frame, but instead that event frames learned in the L1 are not preempted efficiently in L2 production, leading to the production of unidiomatic expression in the L2.

Goldberg (2019, p. 142) explicitly rejects the Whorfian idea that nonlinguistic cognition is influenced by linguistic structures in the L1. Her reasoning for this is the following: Since Portuguese speakers have a single term for lemons and limes (*limão*) and that the third person pronoun in Farsi (/u/) refers to both the male and female gender, it causes speakers of these languages to produce errors when retrieving the intended form in their

L2 English. At the same time, it is fair to assume that speakers of both languages are able to distinguish lemons from limes and men from women. Rejecting any claim that language should constrain general cognition and perception, she proceeds to define the conceptual structures under discussion as representations fixed by categories of language, i.e., representations underlying language comprehension and production. In this context, she appeals to Slobin's (1996) *thinking for speaking hypothesis*, which states that during language production, the conceptualization of spatial relations takes place in accordance with language-specific encoding.

By contrast, Stutterheim, Lambert, and Gerwien (2021; p. 8) note that the assumption of language-specific conceptualization patterns (cf. Slobin, 1996) leaves open the option whether the conceptualization of events in nonlinguistic contexts is based on a universal representation or not (i.e., a representation not specific to the means of encoding in a specific language). It is this type of cognition which is under scrutiny by Stutterheim et al.: Whereas Goldberg focuses on the selection of appropriate expression among competing forms in language production, Stutterheim et al. considers the mechanisms involved in retrieving cognitive representations underlying language production. This means that, although the conceptual representation assumed by Goldberg (the semantic side of a construction) corresponds to the event frame assumed by Stutterheim et al., Goldberg and Stutterheim et al. describe two different mechanisms involved in conceptual representation: One mechanism is responsible for event conceptualization that can serve as basis for linguistic encoding, and since a certain event concept is routinely activated in L1 production, the same concept tends to be activated in L2 production (cf. Stutterheim et al., 2021). At the same time, a different mechanism can be responsible for preempting competing linguistic representations (cf. Goldberg, 2019).

Thus, both mechanisms can be related to the idea that processing language, specifically when two languages (L1 and L2) are competing, involves both the activation of conceptual representations (which can be influenced by entrenched routines of conceptualization in L1) and the inhibition of unintended representations (which can be influenced by transfer effects across languages) (see also research suggesting that in code-switching,

one language is not only "switched on", but another is also "switched off"; Blanco-Elorrieta, Emmorey & Pylkännen, 2018).

However, Stutterheim et al.'s study suggests that language-specific effects are not only to be found in linguistic contexts, but that general cognition is influenced by language. That is, their research suggests that the structures posited by Goldberg must have an underlying conceptualization mechanism that is language-specific, leading to the activation of language-specific event schemata which further triggers the activation of the corresponding linguistic forms.

It is assumed here that Goldberg's assumptions of universality and Stutterheim et al.'s assumptions about language-specificity do not constitute a conflict. To understand why the two approaches are not mutually exclusive, it is crucial to distinguish between, on the one hand, the fixed structures that are stably represented in long-term memory, and, on the other hand, the cognitive mechanisms involved in retrieving and conceptualizing a representation in short-term memory. Arguably, whereas Goldberg's rejection of language-specific concepts refers to representations in long-term memory, Stutterheim et al.'s research makes no claim about any language-specificity of this particular form of representation. Instead, Stutterheim et al.'s research suggests that it is in the pre-linguistic conceptualization, i.e., during the retrieval of representations in short-term memory, that language-specific routines influence cognitive processing. Importantly, this leaves open whether any representations in long-term memory are influenced by the means of a language in the Whorfian sense (see Goldberg's criticism above).

10.5.1.4 The compatibility of the approaches

One might object that the findings presented in Stutterheim, Lambert, and Gerwien (2021) may be exhaustively explained by the principles of coverage and statistical preemption (cf. Goldberg, 2019), rendering any considerations of the role language-specific conceptualization redundant. In this alternative explanation of the data, implicit processing mechanisms in L2 speakers of German and English would show difficulties in preempting the use of the locative event framing when retrieving a manner verb, as this framing has

been learned in their L1 French and causes interference. This would mean that the event frame is not necessarily activated per routine, but instead that its activation is not inhibited, since the negative evidence that this event frame is not used in the L2 is being obscured by knowledge of the L1.

Although this cannot be ruled out, there are reasons that it may seem implausible solely attributing this explanation to the data: First, the high amount of event framing native to their L1 that French speakers produced in the study suggests that it is more than occasional unidiomatic use caused by an inefficient preemption, but that the event frame typically used the L1 indeed is automatically activated in L2 production. Second, other experimental findings support this claim by showing that French speakers segment events according to the propositional structure of their language in nonlinguistic contexts (compared with German speakers), i.e., when experimental participants are not describing events, but instead just pressing a button to mark that a new event is taking place in a dynamic visual scene (i.e., while performing an event segmentation task) (Gerwien & Stutterheim, 2018).

Conversely, it cannot be ruled out either that language-specific routines in conceptualization or in the retrieval of linguistic forms take place, making the principles laid out by Goldberg obsolete. In this explanation, the automatic activation of event frames or syntactic structures (as the DO or PO) would suffice to explain improper L2 use.

However, it is important to keep in mind at what level of linguistic encoding this interference takes place. Whereas Stutterheim et al.'s study focuses on the pre-linguistic conceptualization of motion events, Goldberg focuses on the preemption of competing, unidiomatic linguistic forms potentially describing one and the same conceptualized event.

Hence, this is a matter of further empirical research to investigate which of the above-presented approaches account for asymmetries in L1 and L2 acquisition, and, if both, to which degree they each contribute to ungrammatical and unidiomatic expressions in L2 production.

10.5.2 Multilayered concepts in language production

As we have seen in the theoretical analysis and experimental testing, the DO construction in the English language can be associated with an abstract transfer-of-possession meaning across instances containing different verbs and referent types. These findings are consistent with a Construction Grammar approach to language, according to which abstract syntactic structures, and not just lexical verbs, carry meaning. However, as we have also seen evidence of in chapter 4.2.4.1, the PO construction is not always associated with one particular abstract meaning across instantiations, contra predictions from the Construction Grammar framework that the PO construction should associate caused motion meaning across instances. Evidence for this is that when the PO construction is instantiated with the verb *give*, the construction is associated not with caused motion, but with caused possession. It is, however, central to the robustness and generalizability of a Construction Grammar account to explain how PO construction is supposed to be associated with caused motion, while the frequent implementation with the verb *give* overwrites this meaning with a caused possession meaning.

This challenge can be met by taking a closer look at what exactly constitutes a cognitive representation corresponding to the meaning of the PO construction. As we have seen in the theoretical part of the dissertation, there are reasons to assume that event representations are multilayered, i.e., that they can contain multiple meanings at the same time. In the context of the cognitive representation of the PO construction, this can be narrowed down to the idea that the target of transfer is not only a possessor, but that this possessor is additionally understood as a spatial goal. In psycholinguistic terms, this means that activation of both of these meaning components (possessor and spatial goal) could underlie the choice of this syntactic form. The basic idea can be captured by the principles that Van Valin and LaPolla (1997) use to describe what we call here, based on Löbner (2021), multilayered concepts: an abstract conceptual schema is instantiated by more specific lexical concepts (see chapter 4.3.1; see chapter 4.3.2 for an application of this principle to Construction Grammar). However, at this point, theoretical analysis meets its limits and experimental investigation is required (but see Ziegler, Snedeker & Wittenberg, 2018).

11 Conclusion

This work shed light on aspects of the complex cognitive structures that underlie our mental ability to represent and verbalize actions and events. The description of these cognitive phenomena was primarily approached using two theories, one of them focusing on their multilayeredness and the other on their temporal complexity. It was shown that ideas of multilayeredness and temporality are interrelated. In addition, it was shown that these approaches to understanding cognitive structure can be brought together with ideas about how meaning is constituted in language (at the level of both sentence form and verb meaning).

The methodological part of the dissertation addressed a central aspect of these considerations. In two experiments, the temporal component of language comprehension was explored, as well as the relationship between event semantics and syntactic sentence structures. The results show that (i) native English listeners use aspectual morphosyntax to anticipate and integrate event participants in transfer-of-possession events (expressed by the verbs *hand* and *pass*); (ii) that listeners are able to analyze aspect information marked on a novel verb and simultaneously analyze sentence structure (cf. Construction Grammar) to infer sentence meaning and modulate its temporal properties.

Altogether, the theoretical analyses and experimental findings contribute to an understanding of actions and events that goes beyond a simple semantic categorization such as thematic structures or event structures, but which considers a fine-grained approach to the underlying complex cognitive structures. This opens up the possibility for future research to more precisely investigate the structures that are mentally instantiated during language processing, both at a linguistic and at a pre-linguistic level.

12 References

Altmann, G. T. M., & Ekves, Z. (2019). Events as intersecting object histories: A new theory of event representation. *Psychological Review, 126(6)*, 817-840.

Altmann, G. T. M., & Kamide, J. (2004). Now You See It, Now You Don't: Mediating the Mapping between Language and the Visual World. *The interface of language, vision, and action: Eye movements and the visual world*, 347-386. Psychology Press.

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition, 73*, 247-264. https://doi.org/10.1016/S0010-0277(99)00059-1

Altmann, G. T. M., & Kamide, Y. (2009). Discourse-mediation of the mapping between language and the visual world: Eye movements and mental representation. *Cognition, 111(1)*, 55-71. https://doi.org/10.1016/j.cognition.2008.12.005

Arnold, J. (1996). The Effect of Thematic Roles on Pronoun Use and Frequency of Reference Continuation. *DISCOURSE PROCESSES, 31*, 137-162. https://doi.org/10.1207/S15326950DP3102_02

Asher, N. (2011). Lexical Meaning in Context: A Web of Words. Cambridge University Press. https://doi.org/DOI: 10.1017/CBO9780511793936

Austin, J. L. (1962). How to do things with words (Vol. 75). Clarendon Press.

Bartlett, F. C. (1932). Remembering: A study in experimental and social psychology. Cambridge University Press.

Baumann, M., Pappert, S., & Pechmann, T. (2021). Evidence against lexicalist or configurational approaches to structural encoding in sentence production. *A. Artemis & V. Elisabeth Sophia Maria (Eds.), The Syntax of Argument Structure, 33-68*. De Gruyter. https://doi.org/doi:10.1515/9783110757255-003

Beavers, J., Koontz-Garboden, A. (2020). The Roots of Verbal Meaning. (Oxford University Press)

Beck, Sigrid & Kyle Johnson. 2004. Double objects again. *Linguistic Inquiry 35*: 97-124.

Beebee, H., Hitchcock, C., & Menzies, P. (2009). The Oxford Handbook of Causation. Oxford University Press UK.

Bencini, G., & Goldberg, A. (2000). The Contribution of Argument Structure Constructions to Sentence Meaning. *Journal of Memory and Language, 43,* 640-651. https://doi.org/10.1006/jmla.2000.2757

Bencini, G. M. L. (2013). 379 Psycholinguistics. *T. Hoffmann & G. Trousdale (Eds.), The Oxford Handbook of Construction Grammar*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195396683.013.0021

Bennett, J. (2002). What events are. *R. M. Gale (Ed.), The Blackwell Guide to Metaphysics*. Blackwell.

Bergen, B., & Wheeler, K. (2010). Grammatical aspect and mental simulation. *Brain Lang, 112(3)*, 150-158. https://doi.org/10.1016/j.bandl.2009.07.002

Biber, D. (1986). Strategies of Discourse Comprehension. *Language, 62(3)*, 664-668. https://doi.org/10.2307/415483

Bierwisch, M. (1982). Formal and Lexical Semantics. *Linguistische Berichte 80*: 3-17. Wiesbaden

Blanco-Elorrieta, E., Emmorey, K., & Pylkkänen, L. (2018). Language switching decomposed through MEG and evidence from bimodal bilinguals. *PNAS Proceedings of the National Academy of Sciences of the United States of America, 115(39)*, 9708-9713. https://doi.org/10.1073/pnas.1809779115

Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive Psychology, 18(3)*, 355-387. https://doi.org/10.1016/0010-0285(86)90004-6

Bock, K., & Loebell, H. (1990). Framing sentences. *Cognition, 35*, 1-39. https://doi.org/10.1016/0010-0277(90)90035-I

Branigan, H. (2007), Syntactic Priming. *Language and Linguistics Compass, 1*: 1-16. https://doi.org/10.1111/j.1749-818X.2006.00001.x

Branigan, H. P., & Pickering, M. J. (2017). An experimental approach to linguistic representation. *Behav Brain Sci, 40*, e282. https://doi.org/10.1017/s0140525x16002028

Bücking, S., & Maienborn, C. (2019). Coercion by modification — The adaptive capacities of event-sensitive ad-nominal modifiers. *Semantics and Pragmatics, 12*, 1-39. https://doi.org/10.3765/sp.12.9

Bürkner, P.-C. (2017). brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software, 80(1)*, 1–28. https://doi.org/10.18637/jss.v080.i01

Bürkner, P-C.. (2018). The R Journal. 10:1, 395-411.

Cai, Z. G., Pickering, M. J., & Branigan, H. P. (2012). Mapping concepts to syntax: Evidence from structural priming in Mandarin Chinese. *Journal of Memory and Language, 66*(4), 833–849. https://doi.org/10.1016/j.jml.2012.03.009

Casati, R., & Varzi, A. C. (2020). Events. *E. N. Zalta (Ed.), Stanford Encyclopedia of Philosophy*.

Chomsky, N. (1965). Aspects of the Theory of Syntax (50 ed.). The MIT Press. http://www.jstor.org/stable/j.ctt17kk81z

Chomsky, N. (1986). Knowledge of Language: Its Nature, Origin, and Use. Bloomsbury Academic. https://books.google.dk/books?id=b0VZPtZDL8kC

Chomsky, N. (1993). Lectures on Government and Binding. De Gruyter Mouton. https://doi.org/doi:10.1515/9783110884166

Comrie, B. (1976). Aspect. Cambridge: Cambridge University Press.

Cook, Roy, "Frege's Logic", *The Stanford Encyclopedia of Philosophy (Spring 2023 Edition), Edward N. Zalta & Uri Nodelman (eds.)*. <https://plato.stanford.edu/archives/spr2023/entries/frege-logic/>

Diessel, H. (2013). 346347 Construction Grammar and First Language Acquisition. *T. Hoffmann & G. Trousdale (Eds.), The Oxford Handbook of Construction Grammar*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195396683.013.0019

Diessel, H. (2015). 14. Usage-based construction grammar. *D. Ewa & D. Dagmar (Eds.), Handbook of Cognitive Linguistics,* 296-322. De Gruyter Mouton. https://doi.org/doi:10.1515/9783110292022-015

Dobel, C., Gumnior, H., Bölte, J., & Zwitserlood, P. (2007). Describing scenes hardly seen. *Acta Psychol (Amst), 125(2),* 129-143. https://doi.org/10.1016/j.actpsy.2006.07.004

Dölling, J. (2003). Flexibility in adverbal modification: Reinterpretation as contextual enrichment. *L. Ewald, M. Claudia, & F.-H. Cathrine (Eds.), Modifying Adjuncts, 511-552*. De Gruyter Mouton. https://doi.org/doi:10.1515/9783110894646.511

Dowty, D. (1989). On the semantic content of the notion of "thematic role". *G. Chiercha, B. H. Partee, & R. Turner (Eds.), Properties, types and meaning, 69-129*. Dordrecht: Kluwer. doi:10.1007/978-94-009-2723-0_3

Ellis, N. C. (2002). Frequency Effects in Language Processing: A Review with Implications for Theories of Implicit and Explicit Language Acquisition. *Studies in Second Language Acquisition, 24(2)*, 143-188. https://doi.org/10.1017/S0272263102002024

Elman, J. L., Kehler, A., & Rohde, H. (2006). Event Structure and Discourse Coherence Biases in Pronoun Interpretation. *Proceedings of the Annual Meeting of the Cognitive Science Society, 28*.

Evans, V. (2010). Figurative language understanding in LCCM Theory. *Cognitive Linguistics, 21(4)*, 601-662. https://doi.org/doi:10.1515/cogl.2010.020

Evans, V., & Green, M. (2006). Cognitive linguistics: An introduction. Lawrence Erlbaum Associates Publishers.

Fauconnier, G., & Turner, M. (2002). The way we think: Conceptual blending and the mind's hidden complexities. Basic Books.

Feller, D., Eerland, A., Ferretti, T., & Magliano, J. (2019). Aspect and Narrative Event Segmentation. Collabra*: Psychology, 5.* https://doi.org/10.1525/collabra.182

Ferretti, T. R., Kutas, M., & McRae, K. (2007). Verb aspect and the activation of event knowledge. *J Exp Psychol Learn Mem Cogn, 33(1)*, 182-196. https://doi.org/10.1037/0278-7393.33.1.182

Ferretti, T. R., McRae, K., & Hatherell, A. (2001). Integrating Verbs, Situation Schemas, and Thematic Role Concepts. *Journal of Memory and Language, 44(4)*, 516-547. https://doi.org/https://doi.org/10.1006/jmla.2000.2728

Ferretti, T. R., Rohde, H., Kehler, A., & Crutchley, M. (2009). Verb aspect, event structure, and coreferential processing. *Journal of Memory and Language, 61*, 191-205. https://doi.org/10.1016/j.jml.2009.04.001

Fillmore, C. J. (1968). The case for case. *E. Bach & R. Harms (Eds.), Universals in Linguistic Theory.* Holt, Rinehart, and Winston.

Flecken, M., & Gerwien, J. (2013). Grammatical Aspect influences Event Duration Estimations: Evidence from Dutch. *Proceedings of the Annual Meeting of the Cognitive Science Society, 35.* https://escholarship.org/uc/item/4f50f5h0

Frankland, S. M., & Greene, J. D. (2020). Concepts and Compositionality: In Search of the Brain's Language of Thought. *Annual Review of Psychology, 71(1)*, 273-303. https://doi.org/10.1146/annurev-psych-122216-011829

Galton, A., & Mizoguchi, R. (2009). The water falls but the waterfall does not fall: New perspectives on objects, processes and events. *Applied ontology, 4(2)*, 71-107.

Gawron, J. M. (1986). Situations and prepositions. *Linguistics and Philosophy, 9(3)*, 327 - 382.

Gennari, S., & Poeppel, D. (2003). Processing correlates of lexical semantic complexity. *Cognition, 89(1)*, B27-B41. https://doi.org/10.1016/S0010-0277(03)00069-6

Gerwien, J. (2011). A psycholinguistic approach to AT-structure analysis. *Sprachliche Variationen, Varietäten und Kontexte. Festschrift für Rainer Dietrich, Katharina Spalek, Juliane Domke (ed)*, Stauffenburg Festschriften.

Gerwien, J., & Flecken, M. (2015). There is no prime for time: The missing link between form and concept of progressive aspect in L2 production. *International Journal of Bilingual Education and Bilingualism, 18(5)*, 561-587. https://doi.org/10.1080/13670050.2015.1027144

Gerwien, J., & von Stutterheim, C. (2018). Event segmentation: Cross-linguistic differences in verbal and non-verbal tasks. Cognition, 180, 225-237. https://doi.org/10.1016/j.cognition.2018.07.008

Gerwien, J., & von Stutterheim, C. (2022). Conceptual Blending Across Ontological Domains—References to Time and Space in Motion Events by Tunisian Arabic Speakers of L2 German. *Frontiers in Communication, 7*, 856805.

Gerwien, J., Marberg, I., & Nicolaisen, K. (*in press*). What are Events?. *Gerwien, J., Marberg, I., & Nicolaisen, K. (ed.), Die kognitive Perspektive. Wie Menschen über die Welt sprechen.* Heidelberg University Press.

Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review, 9(3)*, 558-565. https://doi.org/10.3758/BF03196313

Goldberg, A. (1995). Constructions. A Construction Grammar Approach to Argument Structure. Chicago, IL: University of Chicago Press.

Goldberg, A. (2006). Constructions at Work: The Nature of Generalization in Language. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199268511.001.0001

Goldberg, A. (2019). Explain Me This. Princeton University Press. https://doi.org/doi:10.1515/9780691183954

Goldman, A. I. (1970). A Theory of Human Action. Princeton University Press.

Goller, F., Choi, S., Hong, U., & Ansorge, U. (2020). Whereof one cannot speak: How language and capture of visual attention interact. *Cognition, 194*, 104023. https://doi.org/https://doi.org/10.1016/j.cognition.2019.104023

Goodman, N. (1951). The structure of appearance. Harvard University Press.

Green, G. M. (1974). Semantics and syntactic regularity. Indiana University Press Bloomington.

Griffin, Z., & Bock, K. (2000). What the Eyes Say About Speaking. *Psychological science, 11*, 274-279. https://doi.org/10.1111/1467-9280.00255

Gropen, J., Pinker, S., Hollander, M., Goldberg, R., & Wilson, R. (1989). The Learnability and Acquisition of the Dative Alternation in English. *Language, 65(2)*, 203-257. https://doi.org/10.2307/415332

Gruber, J. (1965). Studies in Lexical Relations. Doctoral dissertation, MIT.

Grüter, T., Takeda, A., Rohde, H., & Schafer, A. J. (2018). Intersentential coreference expectations reflect mental models of events. *Cognition, 177*, 172-176. https://doi.org/10.1016/j.cognition.2018.04.015

Hafri, A., Papafragou, A., & Trueswell, J. C. (2013). Getting the gist of events: recognition of two-participant actions from brief displays. *J Exp Psychol Gen, 142(3)*, 880-905. https://doi.org/10.1037/a0030045

Harley, H. (2003). Possession and the Double Object Construction. *Linguistic Variation Yearbook*, 2, 31-70. https://doi.org/10.1075/livy.2.04har

Hartshorne, J. K., & Wittenberg, E. V. A. (2018). Gabriel Radvansky and Jeff Zacks, Event Cognition. Oxford University Press, 2014. Pp. 288. ISBN: 978-0-1998-9813-8 (hardback). *Language and Cognition, 10(2)*, 382-389. https://doi.org/10.1017/langcog.2017.28

Herweg, M. (1989): Ansätze zu einer semantischen Beschreibung topologischer Präpositionen. *Ch. Habel, M. Herweg, K. Rehkämper (Hrsg.): Raumkonzepte in Verstehensprozessen.* Niemeyer: Tübingen. 99–127.

Hindy, N. C., Altmann, G. T. M., Kalenik, E., & Thompson-Schill, S. L. (2012). The effect of object state-changes on event processing: Do objects compete with themselves? *The Journal of Neuroscience*, 32, 5795-5803. https://doi.org/10.1523/JNEUROSCI.6294-11.2012

Hoffmann, T., & Trousdale, G. (2013). The Oxford Handbook of Construction Grammar. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195396683.001.0001

Horchak, O. V., & Garrido, M. V. (2021). Dropping bowling balls on tomatoes: Representations of object state-changes during sentence processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 47(5)*, 838-857. https://doi.org/10.1037/xlm0000980

Hovav, M. R., & Levin, B. (2008). The English dative alternation: The case for verb sensitivity. *Journal of Linguistics, 44*, 129 - 167.

Hume, D. (1748). An Enquiry Concerning Human Understanding. Oxford: Clarendon Press, 2006.

Johnson-Laird, P. N. (1983). Mental Models. Towards a Cognitive Science of Language, Inference and Consciousness. Cambridge, UK: Cambridge University Press.

Juliano, C., & Tanenhaus, M. K. (1994). A constraint-based lexicalist account of the subject/object attachment preference. *Journal of Psycholinguistic Research, 23(6)*, 459-471. https://doi.org/10.1007/BF02146685

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87(4)*, 329-354. https://doi.org/10.1037/0033-295X.87.4.329

Kamide, Y., Altmann, G., & Haywood, S. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language, 49*, 133-156. https://doi.org/10.1016/S0749-596X(03)00023-8

Kang, X., Eerland, A., Joergensen, G. H., Zwaan, R. A., & Altmann, G. T. M. (2020). The influence of state change on object representations in language comprehension. *Memory & Cognition, 48(3)*, 390-399. https://doi.org/10.3758/s13421-019-00977-7

Kang, X., Joergensen, G. H., & Altmann, G. T. M. (2020). The activation of object-state representations during online language comprehension. *Acta Psychologica, 210*. https://doi.org/10.1016/j.actpsy.2020.103162

Kaplan, R., & Bresnan, J. (1982). Lexical-Functional Grammar: A Formal System for Grammatical Representation. In (173-281).

Kaschak, M. P., & Glenberg, A. M. (2000). Constructing meaning: The role of affordances and grammatical constructions in sentence comprehension. *Journal of Memory and Language, 43*, 508-529. https://doi.org/10.1006/jmla.2000.2705

Kay, P., & Michaelis, L. (2012). Constructional Meaning and Compositionality. *Semantics – Interfaces*. De Gruyter Mouton, (pp. 2271-2296). https://doi.org/10.1515/9783110589849-009

Kehler, A., Kertz, L., Rohde, H., & Elman, J. L. (2008). Coherence and Coreference Revisited. *J Semant, 25(1)*, 1-44. https://doi.org/10.1093/jos/ffm018

Klein, W. (1994). Time in Language. Routledge. https://doi.org/https://doi.org/10.4324/978131500380

Klein, W. (1999). Wie sich das deutsche Perfekt zusammensetzt. *Zeitschrift für Literaturwissenschaft und Linguistik, 113, 52-85.* Metzler, Stuttgart.

Klein, W. (2002). The argument-time structure of recipient constructions in German. *Issues in Formal German(ic) Typology*, 141-178. https://doi.org/https://doi.org/10.1075/la.45.08kle

Konopka, A. E., & Bock, K. (2009). Lexical or syntactic control of sentence formulation? Structural generalizations from idiom production. *Cognitive Psychology, 58*, 68-101. https://doi.org/10.1016/j.cogpsych.2008.05.002

Konopka, A. E., & Brown-Schmidt, S. (2014). Message encoding. *M. Goldrick, V. Ferreira, & M. Miozzo (Eds.), The Oxford handbook of language production*, 3-20. Oxford University Press.

Konopka, A. E., Meyer, A., & Forest, T. A. (2018). Planning to speak in L1 and L2. *Cogn Psychol, 102*, 72-104. https://doi.org/10.1016/j.cogpsych.2017.12.003

Krifka, M. (1999). At least some determiners aren't determiners. *K. Turner (Ed.), The Semantics/Pragmatics Interface From Different Points of View*, 1-257. Elsevier.

Krifka, M. (2004). Topic and Focus: A Cross-Linguistic Perspective. Kluwer Academic Publishers.

Lambert, M., von Stutterheim, C., Carroll, M., & Gerwien, J. (2022). Under the surface: A survey of principles of language use in advanced L2 speakers. *Language, Interaction and Acquisition, 13(1)*, 1-28. https://doi.org/https://doi.org/10.1075/lia.21014.lam

Lambrecht, K. (1994). Information structure and sentence form: Topics, focus, and the mental representations of discourse referents. Cambridge University Press. https://doi.org/10.1017/CBO9780511620607

Langacker, R. W. (1987). Foundations of Cognitive Grammar. Vol. 1.: Theoretical Prerequisites. Stanford University Press. https://www.degruyter.com/database/COGBIB/entry/cogbib.7220/html

Langacker, R. W. (2008). Cognitive Grammar: A basic introduction. Oxford: Oxford University Press. Journal of Linguistics, 45.

Lee, S. H.-y., & Kaiser, E. (2022). Mapping language onto mental representations of object locations in transfer-of-possession events: A visual-world study using webcam-based eye-tracking. *Proceedings of the Annual Meeting of the Cognitive Science Society, 44(44)*. https://doi.org/https://escholarship.org/uc/item/40c465mw

Levelt, W. J. M. (1989). Speaking: From intention to articulation. The MIT Press.

Levin, B., & Rappaport Hovav, M. (2005). Argument Realization. Cambridge University Press. https://doi.org/DOI: 10.1017/CBO9780511610479

Levinson, S. (2003). Space in Language and Cognition: Explorations in Cognitive Diversity. *Space in Language and Cognition: Explorations in Cognitive Diversity, Vol. 5.* https://doi.org/10.1017/CBO9780511613609

Lewis, D. (1973). Causation. *Journal of Philosophy, 70(17)*, 556-567.

Löbner, S. (2020). Kaskaden – eine fundamentale Struktur kognitiver Repräsentationen. *Yoshiyuki Muroi (Hg.), Einheit in der Vielfalt? Germanistik zwischen Divergenz und Konvergenz. Asiatische Germanistentagung 2019 in Sapporo*. München: iudicium, 673–692.

Löbner, S. (2021). Cascades. Goldman's Level-Generation, Multilevel Categorization of Action, and Multilevel Verb Semantics. *Löbner, S., Gamerschlag, T., Kalenscher, T., Schrenk, M., Zeevat, H. (eds) Concepts, Frames and Cascades in Semantics, Cognition and Ontology. Language, Cognition, and Mind, vol 7*. Springer, Cham. https://doi.org/10.1007/978-3-030-50200-3_13

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review, 101(4)*, 676-703. https://doi.org/10.1037/0033-295X.101.4.676

Madden, C. J., & Ferretti, T. (2009). Verb aspect and the mental representation of situations. *The Expression of Time*, 217-240.

Madden, C. J., & Zwaan, R. A. (2003). How does verb aspect constrain event representations? *Memory & Cognition, 31*, 663-672. https://doi.org/10.3758/BF03196106

Madden-Lombardi, C., Dominey, P. F., & Ventre-Dominey, J. (2017). Grammatical verb aspect and event roles in sentence processing. *PLoS One, 12(12)*, e0189919. https://doi.org/10.1371/journal.pone.0189919

Magliano, J. P., & Schleich, M. C. (2000). Verb aspect and situation models. *DISCOURSE PROCESSES, 29*, 83-112. https://doi.org/10.1207/S15326950dp2902_1

Magnuson, J. (2019). Fixations in the visual world paradigm: where, when, why? *Journal of Cultural Cognitive Science, 3*, 1-27. https://doi.org/10.1007/s41809-019-00035-3

Majid, A., Bowerman, M., Kita, S., Haun, D. B. M., & Levinson, S. C. (2004). Can language restructure cognition? The case for space. *Trends in Cognitive Sciences, 8*, 108-114. https://doi.org/10.1016/j.tics.2004.01.003

Martin, P., & Bateson, P. P. G. (1993). Measuring behaviour: An introductory guide, 2nd ed. Cambridge University Press. https://doi.org/10.1017/CBO9781139168342

Michaelis, L. (2004). Type shifting in construction grammar: An integrated approach to aspectual coercion. *Cognitive Linguistics, 15*. https://doi.org/10.1515/cogl.2004.001

Michaelis, L. (2017). Meanings of Constructions. *Oxford Research Encyclopedia of Linguistics.* Oxford University Press.

Michaelis, L, & Hsiao, A. M. (2021). Verbing and Linguistic Innovation [Conceptual Analysis]. *Frontiers in Communication, 6*. https://doi.org/10.3389/fcomm.2021.604763

Minsky, M. (1975). A Framework for Representing Knowledge. *The Psychology of Computer Vision, P. Winston (Ed.), McGraw-Hill*.

Misersky, J., Slivac, K., Hagoort, P., & Flecken, M. (2021). The state of the onion: Grammatical aspect modulates object representation during event comprehension. *Cognition, 214*. https://doi.org/10.1016/j.cognition.2021.104744

Oehrle, R. T. (1976). The grammatical status of the English dative alternation. Doctoral thesis, MIT.

Pappert, S., & Pechmann, T. (2014). Priming word order by thematic roles: no evidence for an additional involvement of phrase structure. *Q J Exp Psychol (Hove), 67(11)*, 2260-2278. https://doi.org/10.1080/17470218.2014.918632

Pesetsky, D. M. (1995). Zero Syntax: Experiencers and Cascades. MIT Press.

Pickering, M. J., & Branigan, H. P. (1998). The representation of verbs: Evidence from syntactic priming in language production. *Journal of Memory and Language, 39(4)*, 633-651. https://doi.org/10.1006/jmla.1998.2592

Pinker, S. (1989). Learnability and cognition: The acquisition of argument structure. The MIT Press.

Prystauka, Y. (2018). Comprehending Events on the Fly: Inhibition and Selection During Sentence Processing. University of Connecticut. https://books.google.dk/books?id=42vtzQEACAAJ

Pulvermüller, F. (2013). How neurons make meaning: brain mechanisms for embodied and abstract-symbolic semantics. *Trends Cogn Sci, 17(9)*, 458-470. https://doi.org/10.1016/j.tics.2013.06.004

Pustejovsky, J. (1995). The generative lexicon. The MIT Press.

Pustejovsky, J. (1998). The Semantics of Lexical Underspecification. *Folia Linguistica, 32*. https://doi.org/10.1515/flin.1998.32.3-4.323

Pustejovsky, J., Elisabetta, J. (2016). Integrating Generative Lexicon and Lexical Semantic Resources.

Quine, W. V. O. (1970). Philosophy of logic. Harvard University Press.

Radvansky, G. (2017). Event Segmentation as a Working Memory Process. *Journal of Applied Research in Memory and Cognition, 6*, 121-123. https://doi.org/10.1016/j.jarmac.2017.01.002

Radvansky, G. A., & Zacks, J. M. (2014). Event cognition. Oxford University Press.

Ridge, E. (2013). R. Ellis: The study of second language acquisition. *Per Linguam, 10*. https://doi.org/10.5785/10-1-248

Rohde, H., & Elman, J. (2006). Event structure and discourse coherence biases in pronoun interpretation. *Proceedings of the 28th Annual Conference of the Cognitive Science Society*.

Schank, R. C., & Abelson, R. P. (1977). Scripts, plans, goals and understanding: An inquiry into human knowledge structures. Lawrence Erlbaum.

Slobin, D. I. (1996). From "thought and language" to "thinking for speaking.". Cambridge University Press.

Solomon, S. H., Hindy, N. C., Altmann, G. T., & Thompson-Schill, S. L. (2015). Competition between Mutually Exclusive Object States in Event Comprehension. *J Cogn Neurosci, 27(12)*, 2324-2338. https://doi.org/10.1162/jocn_a_00866

Stefanowitsch, A. (2008). Negative entrenchment: A usage-based approach to negative evidence. *Cognitive Linguistics, 19(3)*.

Stevenson, R. J., Crawley, R. A., & Kleinman, D. (1994). Thematic roles, focus and the representation of events. *Language and Cognitive Processes, 9(4)*, 519-548. https://doi.org/10.1080/01690969408402130

Talmy, L. (2000). Toward a cognitive semantics, Vol. ll: Typology and process in concept structuring. The MIT Press.

Thothathiri, M., & Snedeker, J. (2008). Syntactic priming during language comprehension in three- and four-year-old children. *Journal of Memory and Language, 58(2)*, 188-213. https://doi.org/10.1016/j.jml.2007.06.012

Tomasello, M. (1992). The social bases of language acquisition. *Social Development, 1(1)*, 67-87. https://doi.org/10.1111/j.1467-9507.1992.tb00135.x

Tomasello, M. (2000). First steps toward a usage-based theory of language acquisition. *Cognitive Linguistics, 11*, 61-82. https://doi.org/10.1515/cogl.2001.012

Turner, M., & Fauconnier, G. (2002). The Way We Think: Conceptual Blending And The Mind's Hidden Complexities. Basic Books.

Valin, R., & LaPolla Faha, R. (1998). Syntax: Structure, Meaning, and Function. https://doi.org/10.1017/CBO9781139166799

Stutterheim, C., Andermann, M., Carroll, M., Flecken, M., & Schmiedtová, B. (2012). How grammaticized concepts shape event conceptualization in language production: Insights from linguistic analysis, eye tracking data, and memory performance. *Linguistics 50(4)*, 833-867. https://doi.org/doi:10.1515/ling-2012-0026

Stutterheim, C., Lambert, M., & Gerwien, J. (2021). Limitations on the role of frequency in L2 acquisition. *Language and Cognition, 13*, 1-31. https://doi.org/10.1017/langcog.2021.5

Wechsler, S. (1995). The semantic basis of argument structure. University of Chicago Press.

Welke, K. (2019). Konstruktionsgrammatik des Deutschen. Ein sprachgebrauchsbezogener Ansatz. Berlin, Boston: De Gruyter.

Wheeldon, L. R., & Konopka, A. (2023). Grammatical Encoding for Speech Production. Cambridge University Press. https://doi.org/DOI: 10.1017/9781009264518

Wilson, F., Papafraou, A., Bunger, A., & Trueswell, J. (2011). Rapid extraction of event participants in caused motion events. *Proceedings of the Annual Meeting of the Cognitive Science Society, 33*.

Wolfgang, K., & Ping, L. (2009). The Expression of Time. De Gruyter Mouton. https://doi.org/doi:10.1515/9783110199031

Wulff, S., & Ellis, N. (2018). Usage-based approaches to second language acquisition. *Bilingual Cognition and Language: The state of the science across its subfields*, 37-56. https://doi.org/10.1075/sibil.54.03wul

Zacks, J., & Tversky, B. (2001). Event structure in perception and conception. *Psychological bulletin, 127*, 3-21. https://doi.org/10.1037//0033-2909.127.1.3

Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event perception: a mind-brain perspective. *Psychological bulletin,133(2)*, 273-293. https://doi.org/10.1037/0033-2909.133.2.273

Zacks, J. M., & Tversky, B. (2001). Event structure in perception and conception. *Psychological bulletin, 127(1)*, 3-21. https://doi.org/10.1037/0033-2909.127.1.3

Ziegler, J., Bencini, G., Goldberg, A., & Snedeker, J. (2019). How abstract is syntax? Evidence from structural priming. *Cognition, 193*, 104045. https://doi.org/https://doi.org/10.1016/j.cognition.2019.104045

Ziegler, J., Snedeker, J., & Wittenberg, E. (2018). Event Structures Drive Semantic Structural Priming, Not Thematic Roles: Evidence From Idioms and Light Verbs. *Cogn Sci, 42(8)*, 2918-2949. https://doi.org/10.1111/cogs.12687

Appendix A

Linguistic Stimuli (Experiment 1 and 2)

1. This is Ms. Pink and Ms. Green.
2. This is Ms. Green and Ms. Pink.
3. This is Mr. Blue and Mr. Yellow.
4. This is Mr. Yellow and Mr. Blue.
5. This is Ms. Blue and Ms. Yellow.
6. This is Ms. Yellow and Ms. Blue.
7. This is Mr. Pink and Mr. Green.
8. This is Mr. Green and Mr. Pink.

Critical items (n = 20)

1. This is a rubber glove and a bottle with chemicals.
   Ms. Pink is handing Ms. Green the chemicals.
   Ms. Pink is rubber gloving Ms. Green the chemicals.
2. These are some napkins and a cookie.
   Ms. Pink is handing Ms. Green the cookie.
   Ms. Pink is napkining Ms. Green the cookie.
3. This is a birthday balloon attached to a present.
   Mr. Blue is handing Mr. Yellow the present.
   Mr. Blue is birthday ballooning Mr. Yellow the present.
4. This is a salad fork with a tomato on it.
   Mr. Blue is handing Mr. Yellow the tomato.
   Mr. Blue is salad forking Mr. Yellow the tomato.
5. This is a medal on a trophy.
   Ms. Yellow is handing Ms. Blue the trophy.
   Ms. Yellow is trophying Ms. Blue the medal.
6. This is a piggy bank with a dollar bill in it.
   Ms. Yellow is passing Ms. Blue the piggy bank.
   Ms. Yellow is piggy banking Ms. Blue the dollar bill.
7. This is a tomato plant in a flower pot.
   Mr. Green is passing Mr. Pink the tomato plant.
   Mr. Green is flowerpotting Mr. Pink the tomato plant.
8. This is a frying pan with a fried egg on it.
   Mr. Green is passing Mr. Pink the egg.
   Mr. Green is frying panning Mr. Pink the egg.
9. This is a bucket with popcorn in it.
   Ms. Green is passing Ms. Pink the popcorn.
   Ms. Green is bucketing Ms. Pink the popcorn.
10. This is a palette with a paint brush on it.
    Ms. Green is passing Ms. Pink the palette.
    Ms. Green is paletting Ms. Pink the paint brush.
11. This is an ice tong with some ice cubes.
    Mr. Yellow has handed Mr. Blue the ice cubes.
    Mr. Yellow has ice tonged Mr. Blue the ice cubes.
12. This is a clothespin with a sock.
    Mr. Yellow has handed Mr. Blue the sock.
    Mr. Yellow has clothespinned Mr. Blue the sock.
13. This is a piece of cheese with a butter knife in it.
    Ms. Blue has handed Ms. Yellow the cheese.
    Ms. Blue has butter knifed Ms. Yellow the cheese.
14. This is a little bow attached to a candy cane.
    Ms. Blue has handed Ms. Yellow the candy cane.

Ms. Blue has candy caned Ms. Yellow the bow.

15. This is a camera with a leather strap attached to it.
    Mr. Pink has handed Mr. Green the camera.
    Mr. Pink has leather strapped Mr. Green the camera.
16. This is an eggtray with eggs in it.
    Mr. Pink has passed Mr. Green the eggs.
    Mr. Pink has eggtrayed Mr. Green the eggs.
17. This is a little umbrella in a coconut.
    Ms. Pink has passed Ms. Green the coconut.
    Ms. Pink has coconutted Ms. Green the umbrella.
18. This is a pencil and a pencil sharpener.
    Ms. Pink has passed Ms. Green the pencil and the pencil sharpener.
    Ms. Pink has penciled Ms. Green the pencil sharpener.
19. This is a toy mouse on a pillow.
    Mr. Blue has passed Mr. Yellow the mouse.
    Mr. Blue has pillowed Mr. Yellow the mouse.
20. This is a little dining chair with a doll on it.
    Mr. Blue has passed Mr. Yellow the doll.
    Mr. Blue has dining chaired Mr. Yellow the doll.

Filler items (n = 40)

21. This is an oven mitt and a small ceramic pot.
    Ms. Green is lifting up the pot.
    Ms. Green is oven mitting the pot.
22. This is a grapefruit with a spoon in it.
    Ms. Green is eating from the grapefruit.
    Ms. Green is grapefruiting the spoon.
23. This is a dart arrow in a dart board.
    Ms. Yellow is pulling out the dart arrow.
    Ms. Yellow is dart arrowing the dart board.
24. This is a piece of apple with a flag in it.
    Mr. Yellow is removing the flag.
    Mr. Yellow is appleing the flag.
25. This is a candle in a candlestick.
    Ms. Blue is blowing out the candle.
    Ms. Blue is candlesticking the candle.
26. This is a teddy bear with a bow on it.
    Ms. Blue is untying the teddy bear.
    Ms. Blue is bowing the teddy bear. (26)
27. This is a steak on a dinner plate.
    Mr. Pink is cutting the steak.
    Mr. Pink is dinner plating the steak.
28. This is an envelope with a postcard in it.
    Mr. Pink is signing the postcard.
    Mr. Pink is enveloping the postcard.
29. This is a cup on a saucer.
    Ms. Pink is knocking over the cup.
    Ms. Pink is saucering the cup.
30. This is a kitchen scale with an apple on it.
    Ms. Pink is weighing the apple.
    Ms. Pink is kitchen scaling the apple.
31. This is a rose in a vase.
    Mr. Blue has smelled the rose.
    Mr. Blue has vased the rose.
32. This is a suction cup and a glass eye.
    Mr. Blue has picked up the glass eye.
    Mr. Blue has suction cupped the glass eye.
33. This is a letter clip and a letter.
    Ms. Yellow has fastened the letter.
    Ms. Yellow has letter clipped the letter.
34. These are some marshmallows stuck on a cactus.
    Ms. Yellow has picked off the marshmallows.
    Ms. Yellow has cactused the marshmallows.
35. This is a keychain with a key on it.
    Mr. Green has detached the key.
    Mr. Green has keychained the key.
36. These are chopsticks in a box of noodles.
    Mr. Green has stirred the noodles.
    Mr. Green has chopsticked the noodles.
37. This is a coaster with a bottle on it.
    Ms. Green has blown into the bottle.
    Ms. Green has coastered the bottle.
38. This is a credit card in a wallet.
    Ms. Green has tugged in the credit card.
    Ms. Green has walleted the credit card.
39. This is a cutting board with an onion on it.
    Mr. Yellow has peeled the onion.
    Mr. Yellow has cutting boarded the onion.

40. This is an olive on a toothpick.
    Mr. Yellow has snacked the olive.
    Mr. Yellow has toothpicked the olive.
41. This is a fan and a feather.
    Ms. Yellow is laughing at Ms. Blue.
    Ms. Yellow is faning Ms. Blue.
42. This is duster and a piece of chalk.
    Ms. Yellow is talking to Ms. Blue.
    Ms. Yellow is dustering Ms. Blue.
43. This is a pirate hook and a pirate eye patch.
    Mr. Green is thinking about Mr. Pink.
    Mr. Green is pirate hooking Mr. Pink.
44. This is a sausage in a hot dog bun.
    Mr. Green is listening to Mr. Pink.
    Mr. Green is hot dog bunning Mr. Pink.
45. This is a glasses case with glasses.
    Ms. Green has smiled at Ms. Pink.
    Ms. Green has glasses cased Ms. Pink.
46. This is a magic wand and a wizard's hat.
    Ms. Green has laughed at Ms. Pink.
    Ms. Green has magic wanded Ms. Pink.
47. These are some sardines in a tin can.
    Mr. Yellow has talked to Mr. Blue.
    Mr. Yellow has tin canned Mr. Blue.
48. This is a lunchbox with lunch in it.
    Mr. Yellow has thought about Mr. Blue.
    Mr. Yellow has lunchboxed Mr. Blue.
49. This is a birthday candle in a birthday cake.
    Ms. Blue is listening to Ms. Yellow.
    Ms. Blue is birthday caking Ms. Yellow.
50. This is a toy trolley with a flashlight in it.
    Ms. Blue is smiling at Ms. Yellow.
    Ms. Blue is toy trolleying Ms. Yellow.
51. This is a ping pong racket with a ping pong ball on it.
    Mr. Pink is laughing at Mr. Green.
    Mr. Pink is ping pong racketing Mr. Green.
52. This is a plastic bag with carrots in it.
    Mr. Pink is talking to Mr. Green.
    Mr. Pink is plastic bagging Mr. Green.
53. This is a quill pen in an inkwell.
    Ms. Pink has thought about Ms. Green.
    Ms. Pink has inkwelled Ms. Green.
54. This is a cocktail with a slice of lime on it.
    Ms. Pink has listened to Ms. Green.
    Ms. Pink has cocktailed Ms. Green.
55. This is a key in a padlock.
    Mr. Blue has smiled at Ms. Yellow.
    Mr. Blue has padlocked Ms. Yellow.
56. This is a coat hanger with a scarf on it.
    Mr. Blue has laughed at Mr. Yellow.
    Mr. Blue has coat hangered Mr. Yellow.
57. This is a pin cushion with pin needles in it.
    Ms. Yellow is talking to Ms. Blue.
    Ms. Yellow is pin needleing Ms. Blue.
58. This is a magnet with nails on it.
    Ms. Yellow is thinking about Ms. Blue.
    Ms. Yellow is magneting Ms. Blue.
59. This is a toy boat in a small plastic tub.
    Mr. Green is listening to Mr. Pink.
    Mr. Green is plastic tubbing Mr. Pink.
60. This is a pencil case with a coloured pen in it.
    Mr. Green is smiling at Mr. Pink.
    Mr. Green is pencil casing Mr. Pink.

Appendix B