Non-Actual Motion Expressions in Language and Gesture

Linea Brink Andersen

Supervisor: Prof. Jordan Zlatev

Centre for Language and Literature, Lund University
MA in Language and Linguistics, Cognitive Semiotics
SPVR01 Language and Linguistics: Degree Project – Master's (Two Years) Thesis, 30 credits

March, 2020

Abstract

This thesis studies the phenomenon of non-actual motion (NAM) (Blomberg, 2014; Blomberg & Zlatev 2014), and its expression in language and gestures. It has been documented that many languages employ expressions of motion in descriptions of static scenarios (Amagawa, 1997; Blomberg, 2014; Matsumoto, 1996; Rojo & Valenzuela, 2003, 2009; Stosic & Sarda, 2009), here called NAM-expressions (NAM-Es). It has been proposed that NAM-Es have different experiential motivations (Blomberg 2014; Blomberg & Zlatev, 2014; Langacker, 1987; Matsumoto, 1996; Talmy, 2000a), which may, collectively, be called NAM-motivations.

This thesis discusses four possible NAM-motivations: visual scanning, imagination, affordance of motion, and enactive self-motion, the latter for the first time. An empirical study was conducted to distinguish between the influence of three of these NAM-motivations, visual scanning, affordance of motion, and enactive self-motion, on expressions of non-actual motion in speech and gesture.

Thirty-nine native speakers of Swedish described 20 drawings depicting spatially extended objects, which were varied systematically to favor the different NAM-motivations. It was tested how frequently NAM-Es occurred in description of different stimuli-types, and a semantic framework for the analysis of motion expressions in gestures was developed to test whether different NAM-motivations were reflected in gestural NAM-Es.

It was found that all the NAM-motivations that were investigated appear to function as motivations for NAM-expressions, but to varying extents. It was found that affordance of motion was relevant for both *narrow* and (what is here called) *extended* linguistic NAM-Es, but not for gestures. Enactive self-motion was relevant for extended linguistic NAM-Es, and visual scanning for narrow linguistic NAM-Es. The influence of visual scanning and enactive self-motion on gestural expression of NAM is complex, and needs further research.

Keywords: Cognitive semiotics, Fictive motion, Gesture, Grounding, Non-actual motion, Phenomenology, Subjective motion.

Acknowledgements

Here I am, at the end of a long, and windy road (this is the first and last road/travel pun, I promise!). Many excellent people have helped me get here:

First, and foremost, a gigantic thank you to my supervisor, Jordan, who has been so enthusiastic, knowledgeable, and supportive throughout this entire process. You always seemed to know when I needed a push, and when I needed some slack, and here we finally are. Ta(c)k!

Secondly, I owe a huge deal of gratitude to the members of the research project *Phenomenology and Typology of Motion* (PATOM) at Lund University: Jordan Zlatev, Simon Devylder, Joost van de Weijer, Johan Blomberg, and project assistants and collaborators. The data used in this thesis was collected and transcribed within the PATOM project, and a very useful, custom lexicon was developed by its members. Without these resources, there would not have been a lot for me to discuss on these pages. Writing my thesis in close proximity to this project, I have had the pleasure and benefit of many theoretical and methodological discussions with Jordan, Simon, Joost, and Johan, and they have all been helpful in answering any and all of my questions and encouraging this thesis. Thank you all!

Thank you, also, to my co-students at the Master's program in Language and Linguistics, who ensured that SOL was not only a place for learning, but also for fun and friendship. A special shout-out to Sigrid who lend her photographic expertise to capture the gesture-examples sprinkled throughout Chapter 2, and for being an all-around awesome *human bean*.

Last, but absolutely not least, thank you to my family, who always have my back, and to Cody; your love and support is always felt, even across oceans. I love you.

Table of Contents

List of Figures

List of Tables

Abbreviations

ASP	Aspect
COP	Copula
DEF	Definite
DET	Determiner
DSL	Danish Sign Language
E	Extension
I-NAM	Imaginary non-actual motion
IMPERF	Imperfect
INDF	Indefinite
LOC	Locative
M	Motion
MM	Motion-Manner
MoR	Mode of Representation
NAM	Non-actual motion
NAM-E	Non-actual motion expression
N-NAM-E	Narrow non-actual motion expression
P-NAM	Potential non-actual motion
R-NAM	Resultative non-actual motion
S-NAM	Simulated non-actual motion
SP	Shape/Placement
T	Trace
TOP	Topic
X-NAM-E	Extended non-actual motion expression

1 Introduction

There are no "mere" appearances,
and nothing is "just" an appearance.
Appearances are real; they belong to being
(Sokolowski, 2000, p. 15)

In everyday language use, across many (and possibly all) different languages, it is common to use linguistic expressions of motion to describe situations which do not involve any actual, physical motion. Such expressions, as exemplified by sentences (1.1), (1.2), and (1.3), have been called *non-actual motion expressions* (Blomberg, 2014). Sentences such as these, as well as certain gestural expressions produced while describing static scenes, are the topic of this thesis.

(1.1) The road goes through the forest.

(Blomberg, 2014, p. 160)

(1.2) The fence follows the coastline.

(Richardson & Matlock, 2007, p. 130)

(1.3) Insanity runs in my family… It practically gallops!

(Arsenic and Old Lace, as cited in Brandt, 2009, p. 573)

Motion is ubiquitous in our everyday life. We move, and people and things around us move almost perpetually. We seem to have an intuitive understanding of what motion is, but are hard pressed when we need to define it (Blomberg 2014, p. 3). Within research in motion semantics, defining motion is not straightforward either. Definitions are often based on not only the physical phenomenon, but also the linguistic expression thereof (Zlatev, Blomberg, & David, 2010, p. 389).

Before going on to discuss non-actual motion, we therefore need to establish what actual motion is. Here, I adopt the definition presented by Zlatev, Blomberg and David (2010): "The experience of continuous change in the relative position of an object (the figure) against a

background" (p. 393). It is important to notice that this definition hails from a phenomenological tradition, which is interested in the human experience, rather than so-called "objective reality" (Zlatev, Blomberg, & David, 2010, p. 389).

In the sentences given in (1.1), (1.2), and (1.3), verbs of motion are used; yet it is clear that the scenarios described do not involve actual motion: continuous change of the relative position of an object against a background. The road, fence, and insanity are not moving anywhere.

This phenomenon of describing static events through terms of motion has been studied under different terms, such as *fictive motion* (e.g. Richardson & Matlock, 2007; Talmy, 2000a) and *subjective motion* (Langacker, 1987; Matsumoto, 1996). These, however, are problematic for at least two reasons. First, it is not clear whether they are to be taken as different terms for the same phenomenon, or for different, but related, phenomena. Second, the authors do not clearly state whether they refer to non-linguistic processes of consciousness or cognition, or to (linguistic) expressions.

Therefore, I adopt the terminology of Blomberg (2014) and Blomberg and Zlatev (2014) who take a broad approach, including a wide range of different types of expressions, and systematically distinguish these from the (possible) motivations behind them. Linguistic non-actual motion expressions (henceforth, NAM-expressions) are further defined in Section 2.3.2, but for now it is sufficient to state that a sentence is a NAM-expression, if it uses words with motion semantics, above all verbs, to describe static configurations (Blomberg, 2014, p. 149). Concerning NAM-expressions in gesture, which have not previously been studied, the definition is more complex, and is discussed in Section 3.5. Given that previous research has explored the capacity to express *actual* motion in gestures (see for example Gullberg, 2011; Gullberg, Hendriks, & Hickmann, 2008; Hickmann, Hendriks, & Gullberg, 2011; Özyürek & Kita, 1999; Özyürek et al., 2008), this thesis sets out to explore new territory by asking if gestures may also be used for expression of *non-actual* motion.

An important aspect of previous non-actual motion research has been to establish the *motivations* behind this type of linguistic expressions. Given that NAM-expressions occur across the languages of the world, it is reasonable to assume that they are motivated by prelinguistic, cognitive factors. Based on previous literature on subjective and fictive motion as well as

empirical research, Blomberg (2014) proposed three distinct motivations: visual scanning, enactive perception, and imagination. In brief, visual scanning involves the motion of the gaze itself rather than the object of the gaze, enactive perception concerns our experience of objects as potentials for action, while imagination is conscious imagery of objects as moving. In Section 2.3.1, I provide thorough descriptions and discussions of these possible NAM-motivations, as well as a proposal to distinguish enactive perception into two separate motivations.

Previous research on NAM-expression has focused on linguistic data. However, human communication is for the most part *polysemiotic* (Zlatev 2019). For example, when we communicate face-to-face we do not rely purely on the semiotic system of language (and its sub-system of speech), but rather combine this with the system of gesture, as well as other semiotic systems such as facial expressions, and non-linguistic vocalizations (Zlatev, Żywiczyński, & Wacewicz, submitted). This thesis therefore explores the expression of non-actual motion in gesture, and whether they offer insights into the influence of various types of motivation for NAM-expressions. The research questions can be stated as follows:

- **RQ1** What are the possible experiential motivations for NAM-expressions in language and gesture?
- **RQ2** How do these motivations influence the realizations of linguistic and gestural expression of NAM (in Swedish)?
- **RQ3** How do gesture and speech interact in the expression of NAM?

To answer these questions, I explore the speech and gestures produced by Swedish speaking participants describing picture-stimuli, which were designed to systematically vary along parameters that favor specific NAM-motivations. The distribution of NAM-expressions in speech and gestures across different types of stimuli is analyzed, to explore which NAM-experiences may motivate NAM-expressions, and whether they do so to varying degrees. Furthermore, iconic gestures are categorized using a new iconicity-type scheme, to further explore the possible influence of different NAM-motivations when describing different types of static scenes.

The remainder of this thesis progresses as follows. In Chapter 2, I introduce the necessary theoretical background and existing research, before proposing general hypotheses and an exploratory topic. In Chapter 3, the methodology and materials for the empirical study are presented. The results are reported in Chapter 4. Chapter 5 contains a discussion of the findings, and finally, Chapter 6 gives a summary and conclusions, as well as suggestions for future research.

2 Theoretical background

In this chapter, I provide the necessary theoretical background that motivates the hypotheses proposed in Section 2.4. First, I give some essential background about cognitive semiotics and present the general theoretical and methodological framework of this thesis. In Section 2.2, I provide a theoretical definition of gestures, which is employed throughout this thesis, and I evaluate how gestures may be analyzed semantically, before I introduce and discuss the concept of non-actual motion (NAM) in Section 2.3, including both possible NAM-motivations, and previous research on the linguistic expression of NAM in language. This discussion is then extended with a discussion of NAM-expression in gesture in Chapter 3.

2.1 Cognitive semiotics

2.1.1 Methodological triangulation and the cognitive-empirical loop

Cognitive semiotics is a relatively young transdisciplinary academic field, which integrates methods and theories from (primarily) cognitive science, linguistics and semiotics (Zlatev, 2015a), with the general aim of studying meaning-making. As argued by Sonesson (2012), phenomenology as philosophy and methodology (see Section 2.1.2) is an important tool in the process of marrying the humanities with cognitive science. Cognitive semiotics integrates methods from a first-person ("subjective"), second-person ("intersubjective"), and third-person ("objective") perspective, through its (specific kind of) methodological triangulation (see Table 1). In the present thesis, methods from all three perspectives are combined. The first-person perspective provides a basis for the proposal of distinct experiential NAM-motivations on the basis of phenomenological analysis (as further discussed in Section 2.3.1). NAM-expressions in speech and gesture were classified based on systematic intuitions (as further discussed in

Sections 3.4 and 3.5). Gestures were identified, and distinguished from other bodily actions from a second-person perspective, relaying on empathy with the participants to be able to identify communicative intent and representational complexity, as introduced in Section 2.2. Motion verbs were identified with the use of a pre-existing lexicon, meaning that this part of the analysis was detached from the perspective of the author, that is, from a third-person perspective, but see further discussion in section 3.4 about the production of the lexicon. Finally, statistical analysis provides a third-person perspective on the data originating from the first- and second-person analyses.

Table 1 - Methodological Triangulation (adapted from Zlatev, 2015a), along with its application in this thesis

Perspective	Methods	Applications in thesis
First person ("subjective")	Phenomenological analysis Systematic intuitions	Analysis of NAM-motivations Classification of NAM-Es in speech Classification of gestures
Second person ("intersubjective")	Empathy Imaginative projection	Identification of gestures
Third person ("objective")	Detached observation Experimentation	Identification of motion verbs Statistical analysis

The combination of different kinds of methods is one of the strengths of cognitive semiotics, and one that makes it especially apt for studying complex phenomena such as non-actual motion, in experience and expression. Cognitive semiotics similarly aims to balance empirical investigations with conceptual explorations, by employing the circular, perpetual process of the conceptual-empirical loop, as illustrated in Figure 1. When studying any given phenomenon, X, cognitive semiotics asks the question, *What is X?* before venturing in to asking *How does X?* (e.g. function in interaction, emerge in children, evolve in the species, etc.). The answers found to either question in turn influences the other, and the loop may thus repeat indefinitely, each time enriching our understanding of X.

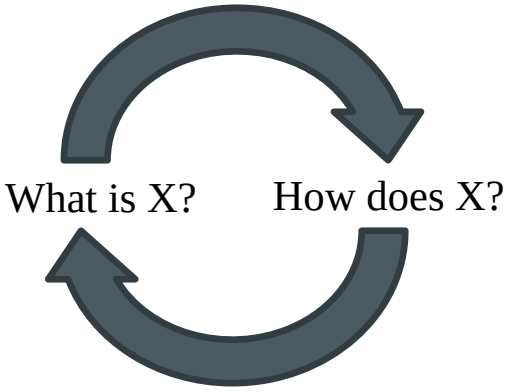


Figure 1 - The conceptual-empirical loop (adapted from Zlatev, 2015a)

2.1.2 Phenomenology

Phenomenology may be characterized as "the careful description of what appears to consciousness precisely in the manner of its appearing" (Moran, 2005, p. 1). It is an important insight in phenomenological philosophy that consciousness is always directed toward something, an intentional object, and this directedness is called *intentionality* (Sokolowski, 2000, p. 8). This focus on intentionality makes phenomenology highly relevant not only to cognitive semiotics in general, but to this thesis in particular, as it aims to explore how our experience of different phenomena may potentially motivate the way we describe them verbally and through gestures. Sokolowski (2000) states that "for phenomenology, intentionality is highly differentiated. There are different kinds of intending, correlating with different kinds of objects" (p. 12). This highlights the fact that experience is not a homogenous phenomenon, and in Section 2.3.1 we will see how this can be applied to various NAM-experiences as possible motivations for NAM-expressions.

2.1.3 Signs, grounds and semiotic systems

Semiotics is sometimes defined as the study of signs (Chandler, 2007) but can more generally be understood as the study of meaning (Sonesson, 1992). Especially for cognitive semiotics, as the transdisciplinary study of (human) meaning-making, sign use is only one particular kind of

significance (Zlatev, 2009, 2018). Nevertheless, signs and their functions are central concepts to cognitive semiotics in general, as well as to this thesis, and a definition is therefore in place. Sign (use) is defined by Zlatev, Żywiczyński and Wacewicz (submitted, p. 10) as an expression (E), used by a Subject (S), to signify an intentional object (O) for (another) S, if and only if:

(a) E and O are linked: in perceiving or enacting E, S conceives of O.

(b) E and O are differentiated: E is qualitatively different from O for S.

(c) The relation is asymmetrical (E -> O).

As is made clear from this definition, sign (use) is better understood as a process, rather than an object, since the sign is not synonymous with the expression, but only exists in usage, i.e. an *expression used by someone to signify something to someone*; that is: to someone else in communication, or to oneself in thought.

The intentional object, which may be an entity, action, event, feeling, etc., is what the sign is about. We may understand the intentional object within the phenomenological tradition as the target of consciousness, that which we are conscious of. It is therefore not necessarily physical, but may well be imagined, remembered, etc. As stated in condition (b) and (c), the expression and object must not only be perceived to be qualitatively different, but this relation must also be asymmetrical, in the sense that one part, the *expression*, is more directly experienced, while the other part, the *object*, is more directly in focus (Sonesson, 2010). Despite being different, there must be some link, or relation, between expression and object, which allows the conception of the object, from the expression. These relations have been identified as the three *semiotic grounds*: iconic, indexical and symbolic (Sonesson, 2010), with roots in Peircian semiotics (Peirce, 1894).

The *iconic ground* is resemblance based, meaning that there is some aspect of similarity between the expression and the object, as in the case of a picture of a tree, or an iconic gesture representing a tree by outlining it in the air. The *indexical ground* is based on contiguity or on part-whole relations, meaning that there is some aspect of proximity in time or space between the expression and the object, e.g. in the case of a footprint, or index-finger pointing. The *symbolic ground* is a relationship of convention between the expression and the object, which can be understood in terms of *common knowledge* (Itkonen, 2008). An example would be the English

word *tree*, or Danish *træ* which both may both be used (within the respective speech community) to refer to a tree.

The semiotic grounds can freely combine in one and the same sign (Jakobson 1965; Andrén, 2010). For example, the Danish Sign Language (DSL) sign for TREE (illustrated in Figure 2), has both an iconic and conventional ground. The sign shows a resemblance between the shape of the forelimb (expression) and the intentional object. However, there are also normative conventions of being well-formed for the sign within DSL, meaning that the sign is also symbolic (McNeill, 2005, p. 8). Despite this possibility for multiple grounds, it is often the case that we may identify one ground to be dominant (Jakobsoni 1965), in the sense of being the primary ground by which the expression and intentional object is perceived to be linked, in a particular context.

Figure 2 - Static representation of the DSL sign for TREE

Signs (or signals, cf. Zlatev et al., submitted) combine to form semiotic systems, consisting of signs of a particular kind, as well as "inter-sign relations, and properties that are in part determined by the medium employed by the particular system" (Stampoulidis, Bolognesi, & Zlatev, 2019, p. 7). The most well-researched semiotic system is language, but others, such as gesture and depiction are also being studied both within and outside the field of cognitive semiotics.

While different semiotic systems may be used in isolation, it is most often the case that they co-occur in interaction, for example the co-occurrence of language and depiction in street art (Stampoulidis, Bolognesi, & Zlatev, 2019), or gestures and language in face-to-face

interaction. Acts of communication that employ more than one semiotic system are acts of polysemiotic communication, or *polysemiosis* (Zlatev, 2019). The semiotic system of gesture requires some extra attention in the present context, to which we turn in the following section.

2.2 Gesture

2.2.1 Defining gestures

Gestures can be defined as bodily "actions that have the features of manifest deliberate expressiveness" (Kendon, 2004, p. 15). This highlights the properties of being communicative and used in interaction, but it does not make explicit what it means for an action to be deliberately expressive. To further define gestures, and distinguish them from other bodily actions, Andrén (2010) and Zlatev (2015b) state that a bodily action is a gesture if it displays the highest degree of (a) communicative explicitness, (b) representational complexity, or (c) both (cf. Table 2)

Table 2 - Three levels of communicative explicitness and representational complexity (adapted from Zlatev, 2015b)

Communicative explicitness	**Representational complexity**
Explicitly other-oriented action: Visible communicative intent	Explicit sign: Expression E *stands for* intentional object X
Action framed by mutual attunement	Typified aspects of meaning: Expression E *counts as* action A
Side effect of co-presence	Situation-specific aspects of meaning

The highest level of communicative explicitness may be understood in the sense of "second-order Gricean communicative intent" (Andrén, 2010, p. 27), where the communicator not only intends for the utterance to express something (first-order intent), but also intends for it to be recognized as such by the audience (second-order intent). A similar concept is that of *intimation*. According to Husserl (1901), intimation is a fundamental feature of the sign use, by

which, apart from the signification of meaning, there is also the announcement of "the speaker's intention to express himself" (Bundgaard, 2010, p. 374). This entails that "[…] the hearer intuitively [*anschaulich*] takes the speaker to be a person who is expressing this or that" (Husserl, 1901, as cited in Bundgaard, 2010, p. 374). In other words, a gesture may be recognized not only for expressing a certain intentional object X, but also for being intended to communicate something by someone.

The highest form of representational complexity implies that the gesture functions as a fully-fledged sign (see Section 2.1.3). Yet, a gesture may also display lower levels of representational complexity, in which the expression does not *stand for*, but rather *count as* a typified action, given that it displays the highest form of communicative explicitness, for example when waving goodbye.

Bodily actions which fulfill these criteria for gestures, are distinguished by a *lower boundary* (see Figure 3) from acts of "body language" such as postures, practical actions performed without communicative intent e.g. itching your nose, and bodily expressions performed with low or no degree of volition, such as blushing (Andrén, 2010, p. 13). Other bodily acts may fulfill the criteria, but may be considered distinct from gestures by an *upper boundary* that separates the bodily expressed signs of a systematic signed language, such as DSL, from gestures (Andrén, 2010).[1] A strict separation between the categories based on the upper and lower boundary, as displayed in Figure 3, is said to be an idealization (Andrén, 2010), as there will always be cases along the borders that may not fall clearly on one or another side of the boundaries.

[1] As discussed by Andrén (2010), these boundaries are idealizations, but as this thesis does not include any further research on signed languages, I will not go into details in the discussions of similarities and differences between gestures and the signs of sign languages

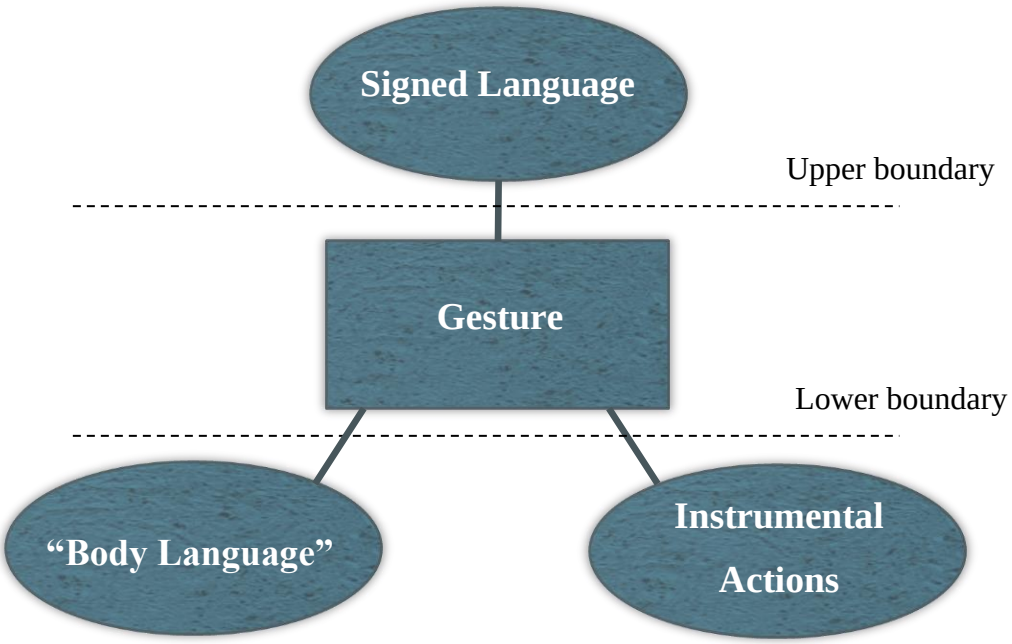


Figure 3 - The upper and lower boundary surrounding gestures (adapted from Andrén, 2010)

Gestures, like other signs, are often subcategorized based on their semiotic ground (see Section 2.1.3), leading to a three-way distinction between iconic gestures, which are similarity based; indexical gestures, which are proximity-based; and emblematic gestures, which are based on conventionality.[2] As for signs in general, gestures may combine features of multiple semiotic grounds, and yet be separated into these three categories based on their predominant ground (Zlatev, 2015b). The gestures most relevant for this thesis are predominantly iconic: "where there is resemblance between the movements of the whole body, or parts of it, and properties of intended actions, objects or whole events" (Zlatev, 2015b, p. 461). When studying iconicity in gestures, as well as in other semiotic systems, it is relevant to consider both *primary iconicity*, where the similarity is such that the interpreter can determine the meaning of the gesture on this

[2] It should be noticed that gesture categorization is not limited to semiotic grounds. Scholars have also studied and categorized gestures with pragmatic functions, for example 'pragmatic gestures' (Kendon, 2004), which can serve a range of functions, like marking illocutionary force, grammatical marking, parsing, etc. (p. 5), or 'beat gestures', which may be used to place emphasis on certain words or phrases in speech (McNeill, 1992, pp. 15-16).

basis, and *secondary iconicity*, where the interpreter needs to first be aware that there is a relation between expression and intentional object, before recognizing the similarity (Sonesson, 1997).

Importantly, iconic gestures may be sub-categorized along different dimensions, as discussed in the following sub-section.

2.2.2 Semantic classification of iconic gestures

Multiple frameworks have been presented for categorizing iconic gestures based on different types of iconic resemblance between the expression and intentional object. In this thesis, I explore whether different kinds of iconic gestures are motivated by different NAM-experiences. This requires the development of a specific set of criteria that may be used to distinguish between different kinds on NAM-expressions in gestures. Such a framework is introduced in Section 3.5, but first it may be helpful to explore existing frameworks for (general) semantic analysis of iconic gestures.

Perhaps the most prominent framework for semantic gesture analysis of iconic gestures is Müller's Modes of Representations (MoR), which distinguishes between four "modes" by which iconic gestures may be produced: *acting*, *molding*, *drawing* and *representing* (Müller, 2014; see also Kendon, 2004). Each mode highlights certain features of the intended object. In the *acting* mode, miming or re-enacting may be used to represent manual activities with and without imaginary objects. In the *molding* mode, the hands mold a "transient" sculpture and thus features such as shape and size of the represented object may be expressed. Something similar may be attained in the *drawing* mode, where the "hand(s) outline(s) the contour or the form of objects or the path of movements in space" (Müller, 2014, p. 1691), which then, in addition to shape and size of objects, can also be used to represent paths of motion. Finally, in the *representing*[3] mode,

[3] The use of the term 'representing' here is challenging, since we may say that all gestures represent an intentional object. It may therefore be more useful to refer to these as "embodying" gestures.

“the hand embodies an object as a whole” (Müller, 2014, p. 1691), and any movements then performed, are representative of the motions of the embodied object.

As can be seen, these modes are theoretically distinct, but they may overlap in actual usage. A possible example would be a DRINKING WATER gesture (as illustrated in Figure 4), in which the speaker forms a horizontal c-shape with one hand, which is then “lifted” up towards the mouth. In that case, we may see the gesture as a whole as *acting* a water-drinking event, but simultaneously the hand *molds* the cup. Müller’s solution to such issues is a conflation of the four types into a binary distinction between *acting* gestures, including enactment of movements and actions with and without objects and *representing* gestures, including representing objects and objects in motion. This grants primacy to the enactment in the example above, although both aspects are (at least theoretically) equally available in perception. This is not to say that the MoR framework is not useful, as it clearly highlights that there are many different ways, or modes, in which a speaker may gesturally represent objects iconically. However, the example illustrates that there are certain challenges to applying the Modes of Representation as discrete “types” in gesture coding.

Figure 4 - Example of DRINKING WATER gesture

A second framework of semantic analysis of “gesture-to-world mappings” (Cartmill, Rissman, Novack, & Goldin-Meadow, 2017) is based on an explicit focus on “the relation between the form of the hand and features of the object in the world that the gesture represents” (p. 7). Cartmill et al. (2017) present three categories of hand-mappings: Hand-as-Hand, Hand-as-

Object, and Hand-as-Neutral. In a gesture mapping *hand-as-hand*, the speaker uses their hands to represent hands[4], for example in a gesture representing climbing a ladder, the speaker could shape the hands as if grabbing rungs one by one with alternating hands, as shown in Figure 5. In a *hand-as-object* gesture, the hand maps on to objects, or parts of them. An example could be using the entire forearm to represent a ladder, leaning diagonally, as shown in Figure 6. The hand-as-object category also includes mappings in which the hand is used to represent some other body-part, for example legs, in a gesture where the index and middle finger are extended and alternatingly moved upwards, to represent legs taking steps up a ladder, as shown in Figure 7. Lastly, *hand-as-neutral* gestures are negatively defined, as gestures in which "the shape of the hand itself [does] not represent any features of the referent" (Cartmill et al., 2017, p. 8). According to the authors, these gestures are typically performed either with the index-finger or with a loose hand with an open flat palm. An example of this could be tracing a line upwards with the index finger while talking about climbing a ladder, as shown in Figure 8.

Figure 5 - Example of CLIMBING LADDER *gesture, performed with hands-as-hands*

[4] In Cartmill et al.'s (2017) framework, only hand-gestures are included, but it is reasonable to extend this category to include gestures produced with other body-parts representing corresponding body-parts. Thus, if a person lifts their feet one-by-one in place to represent walking, this could be considered a feet-as-feet gesture, and would belong to this category.

Figure 6 - Example of LADDER *gesture, performed with hand-as-object*

Figure 7 - Example of CLIMBING LADDER *gesture, performed with hand-as-object*

Figure 8 - Example of CLIMBING LADDER *gesture, performed with hand-as-neutral*

The hand-mapping framework applies more readily to gesture coding than Müller's Modes of Representations, as it isolates a single feature of the gestures. In an analysis of non-actual motion expression in gestures it is important that features are not collapsed implicitly, as each may provide relevant insights. I therefore present a framework for analysis of NAM-

expression in iconic gestures in Section 3.5, that explicitly addresses features of dynamicity, hand-mapping, vectors, and their combinations. But first, I present a description of some key structural properties of gestures.

2.2.3 Segmentation and the expression of meaning in gestures

When analyzing gestures, scholars often distinguish between the expressive parts of the movement, and the parts of the movement that may precede and follow the expressive parts, but are not expressive themselves, following an influential framework presented by Kendon (1980, 2004; see e.g. Green, 2014; Kita, van Gijn, van der Hulst, 1998; McNeill, 1992). The movement is segmented into *preparation*, *stroke* and *retraction*, with possible holds at different stages of movement. The stroke is the part that occurs closest to the apex, the point "of furthest remove from the position of relaxation" (Kendon, 2014, p. 112), and is defined as "the phase of the excursion in which the movement dynamics of 'effort' and 'shape' are manifested with the greatest clarity" (Kendon, 2004, p. 112). The stroke is typically preceded by a preparation phase, consisting of movements placing the limbs in location for the stroke. Following the stroke may be an optional post-stroke hold, where the articulator(s) are sustained in place, and finally the gesture ends with a recovery in which the articulator is relaxed or withdrawn. However, multiple strokes may be strung together, meaning that recovery will not occur until the end of the last stroke. Table 3 gives an overview of Kendon's gesture segments.

The stroke and (potential) post-stroke hold together forms a *nucleus*, which is typically interpreted as the expressive part of the gesture, which "usually bracket[s] a semantically complete phrase of speech" (Kendon, 2004, p. 112). However, the interaction between speech and gesture is not a straight-forward matter. Gestures may be used both in conjunction with speech, and alone (Andrén, 2010). When gestures occur in conjunction with speech, they may either be *complimentary*, meaning that the same information is presented in both gesture and speech, or *supplementary*, meaning that the gesture expresses some aspect of information not included in speech (Iverson & Goldin-Meadow, 2005).

Table 3 - Kendon's (2004) gesture segmentation. (adapted from Andersen, 2018)

Preparation	**Stroke**	**(Post-stroke hold)**	**Recovery**
	Nucleus		

While all gestures are dynamic per definition, the nucleus need not be. According to Kendon (2004), what occurs at the apex may either be the performance of some movement, or a short hold of the articulator(s) in a given position (p. 112). Nuclei consisting solely of a hold at the apex have also been termed stroke-less holds (Kita et al., 1998). This could be imagined in a gesture indicating a person's height by holding a flat hand out into the air to the side of the body, at the height of the person, as exemplified in Figure 9. In this case, the movement towards that point is part of the preparation, while the nucleus consists of the static hold, before recovery starts. However, it would also be possible to perform a gesture where the same height is indicated, but the hand is moved upwards until the height is reached, profiling not only the current height, but also the prior process of growth, as exemplified in Figure 10. In this case, the nucleus itself would be dynamic. The distinction between preparation and nucleus thus relies both on the relative properties of the formal features of the gesture, that is hand-shape, movement pattern, tension of the hand, etc., being more defined during nucleus than preparation and retraction (Kendon, 2004, p. 112), as well as on the semantic content of the gesture in relation to the context of speech and world knowledge. In the example of Figure 9, the movement towards the apex (the preparation) is diagonal, which would be indicative that it is not part of the nucleus, since people do not grow in a diagonal direction, and the hand is not flexed until the last picture, which represents the static nucleus.

Figure 9 - Example of HEIGHT *gesture with static nucleus. Only last picture represents nucleus.*

Figure 10 - Example of HEIGHT *gesture with dynamic nucleus. All three pictures represent nucleus.*

2.3 Non-actual motion theory

The phenomenon of describing static scenarios in terms of motion expressions, as illustrated in the Japanese (2.1), Spanish (2.2), and Thai sentence (2.3), has been documented for a wide range of languages, including English, French, Thai, Swedish, Japanese, Serbian, and Spanish (Amagawa, 1997; Blomberg, 2014; Matsumoto, 1996; Rojo & Valenzuela, 2003, 2009; Stosic & Sarda, 2009), and is potentially universal.

(2.1)	Sono	sanmyaku	wa	nan-boku	ni	hashitte	iru.
	The	mountain.range	TOP	south-north	LOC	run	ASP

'The mountain range runs from north to south.'

(Matsumoto, 1996, p. 199)

(2.2) El valle ascendía lentamente hacia el norte.
The valley ascend.IMPERF slowly towards the north

'The valley ascended slowly towards the north.'

(Rojo & Valenzuela, 2009, p. 249. Gloss added)

(2.3) Mi saphan khaâm heˇo.
COP bridge cross ravine

'A bridge crosses a ravine'

(Blomberg, 2014, p. 195)

As pointed out in Chapter 1, this phenomenon has been described under a number of different terms, such as *fictive motion* (Talmy, 2000a), *subjective motion* (Langacker, 1987), and *abstract motion* (Matlock, 2010). These terms are sometimes used interchangeably (e.g. Matsumoto, 1996), but Blomberg and Zlatev (2014) have pointed out that the range of expressions included in each case differs, and that each is laden with theoretical implications about the cognitive processes assumed to underlie the use of motion expressions to describe static situations. While Talmy (2000a) appeals to a "cognitive bias for dynamism" (p. 101), Langacker (1987) focuses on the dynamic nature of the perceiver's focus, so-called visual scanning, and from a cognitive psychology perspective, Matlock (2010) posits "mental simulation" of motion as the motivation of these expressions.

To address these shortcomings, Blomberg (2014) and Blomberg and Zlatev (2014) introduced the theory of non-actual motion (NAM). This theory has contributed an important conceptual distinction between NAM-expressions (NAM-E), and the experiential motivations that may underlie such expressions, NAM-motivations, which are to be understood as pre-linguistic, cognitive phenomena. The latter had not been clearly distinguished from the linguistic expressions in the previous cognitive linguistic and psychological literature.

In addition, NAM theory (Blomberg, 2014, 2015, 2017; Blomberg & Zlatev, 2014) took a unifying approach, focusing widely on expressions employing motion semantics to describe static configurations (Blomberg, 2014, p. 149). To gain insight into the phenomenon the authors considered both differences and similarities between different types of NAM-expressions, as a way to explore their possible experiential motivations.

As noticed by Blomberg and Zlatev (2014), the various experiential motivations proposed by previous authors are not in fact mutually exclusive, and thus, the theories need not be pitted

against each other. Rather, NAM-theory introduced three separate, but complimentary types of experiences, based on phenomenological reanalysis of the previous theories, which may all be taken to be possible motivations for NAM-expressions: *visual scanning*, *imagination* and *enactive perception*. I describe these below, before proposing a further distinction of enactive perception into two separate experiential phenomena, *motion affordance* and *enactive self-motion*, respectively.

2.3.1 Non-actual motion in experience

2.3.1.1 Visual scanning

The motivation of visual scanning is evoked in Langacker's (1987) notion of subjective motion, or more generally *subjective construal*, which stands in contrast to *objective construal*. According to this distinction, any intentional object may be objectively construed, meaning that it is presented as an explicit, focused object of conception, or subjectively construed, meaning that it presented as an implicit, unselfconscious subject of conception (Langacker, 2006, p. 18). Blomberg and Zlatev (2014) point out that this corresponds to the distinction between intentional object (*noema*) and intentional act (*noesis*), in the phenomenological tradition (Sokolowski, 2000). Human consciousness can either focus on the object itself, or on the act of intending it. Applied to semantics, this mean that we can either focus on the static object (e.g. the slope of the mountain), or the motion of our gaze, as we "scan" it. We may thus understand visual scanning as re-focusing on the noesis rather than the noema. This entails that visual scanning as an experiential motivation for NAM-expressions such as (2.4), is related to the experience of actual motion, underlying the expression in (2.5).

(2.4) The trail rises steeply near the summit.

(Blomberg & Zlatev, 2014, p. 405)

(2.5) The balloon rises.

(Blomberg & Zlatev, 2014, p. 405)

2.3.1.2 Imagination

Imagination as a motivation for NAM-expressions is taken to be a conscious and creative process, by which the speaker is construing a static object "as if" it was performing some movement in the sense of personification of "animalification" (Blomberg & Zlatev, 2014). Therefore, the resulting NAM-expressions can be seen as metaphors. According to Blomberg (2015), NAM-expressions motivated by imagination may be especially rich in information pertaining to the manner of motion, as seen in the examples (2.6) and (2.7):

(2.6) The highway crawls through the city.

(Matlock, 2004, p. 232)

(2.7) The dark velvet ditch creeps by my side.

(T. Transströmer, *April and Silence*)

It is relevant to notice that imagination "does not posit existence" (Blomberg & Zlatev, 2014, p. 408), meaning that in contrast to the perception-based motivations, in imagination, the experience of motion is rather created, not perceived.

The type of NAM-expressions rich in manner information that may be taken to be metaphors have only been studied systematically in English, and it has been shown that linguistic conventions prohibit these types of NAM-expressions in Japanese and Yucatek Maya and appear to have only limited occurrences in Swedish (Blomberg & Zlatev, 2014, p. 409). For these reasons, this thesis does not focus on this motivation in the empirical study described in the following chapters. It would be a matter for future research to ask whether stimuli may be designed to systematically elicit these types of expressions, or if it is the communicative context (e.g. poetic language as suggested by Blomberg and Zlatev, 2014) rather than the described objects, that elicits these types of expressions.

2.3.1.3 Enactive perception: motion affordance and enactive self-motion

The notion of enactive perception, also stemming from the phenomenological tradition, was introduced as a reanalysis of Talmy's (2000a) proposal of fictive motion (Blomberg, 2014). Talmy (2000a) contrasts "fictive" to "factive" motion, a distinction said to be based on two modes of cognition, one which is highly perceptually salient, but not veridical, and one which is

veridical, but has a low degree of perceptual salience. Talmy (2000a) gives an example of a framed picture that hangs skewed such that one of its corners points upwards. The "factive" perception of this situation would be of this as a static object with the shape of a rhomb. In the "fictive" perception, however, we would conceive of this as a skewed square needing to be corrected (Blomberg, 2014, p 157). According to Talmy's idea of a cognitive bias towards dynamism "we attend primarily to change, to that which differs and to that which we can affect" (Blomberg & Zlatev, 2014, p. 402), which gives the "fictive" perception priority over the "factive" one.

However, Talmy's view has a number of theoretical problems, as he argues that the fictive and factive construals are mutually exclusive, which leads to the question of how it is possible to obtain knowledge of the static properties of an object at all. Blomberg and Zlatev (2014) therefore reinterpret Talmy's fictive/factive distinction as that of "two modes of givenness that need not be mutually exclusive or even in conflict with one another" (p. 403). These are, respectively, a pre-reflective, *enactive* mode corresponding to Talmy's fictive mode, in which we perceive representational artifacts, and the actions they invite, and a reflective, *distanced* mode corresponding to Talmy's factive mode, in which we may observe formal features of objects, their location, orientation, shape, material, etc. (Blomberg & Zlatev, 2014, p. 403).

The pre-reflective, enactive mode may be understood in terms of Gibsonian affordances, as elaborated by Blomberg (2014, 2015). According to ecological psychology (Gibson, 1979), as well as phenomenology, perception of an object does not only pertain to its physical properties, but also to the possible actions that it "invites":

> To visually perceive a door handle is not merely to see a protruded part of an otherwise flat object (i.e. a door), but primarily (given previous experiences with doors) as something that allows – or affords – entering and exiting a room or a building. (Blomberg, 2017, p. 211).

The affordance of motion can be taken to be an experiential motivation for NAM-expressions pertaining to traversable objects such as roads, paths, ladders, etc., as it may be argued that to

perceive such objects, is also to perceive them as means for (human) motion.[5] Thus, affordance of motion may motivate NAM-expressions such as (2.8):

(2.8) The road goes through the forest.

(Blomberg, 2014, p. 160)

In previous empirical studies of NAM-motivations, it was hypothesized that the motion-affordance motivation would necessarily be stronger when traversable items are presented from a first-person viewpoint, as opposed to a third-person viewpoint (Blomberg, 2014, 2017; Blomberg & Zlatev, 2014). Blomberg (2017) states that "[m]otion affordance […] is immediately connected to the capacity for self-motion" (p. 212), and Blomberg (2014) predicts that a first-person perspective will encourage re-enactment of self-motion. This argumentation is based on the notion of *kinesthesis* (Husserl, 1973) by which "perception is effectuated by the correlation to possible bodily movements" (Blomberg & Zlatev, 2014, p. 403). When perceiving an object, we have freedom to move, which in turn results in a different vantage point on that object. We can never perceive the full three-dimensional object at once, but we are able to combine all the different vantage points into a unity, and we do so in part due to our own mobility.

However, the perception of three-dimensional objects is made possible, not only through self-motion, but also by *transcendental intersubjectivity* (Zahavi, 2003; Blomberg and Zlatev, 2014). This is the implicit understanding that an object may be observed from any other vantage point, which may at any point be inhabited by another individual, who would then have a different perspective on the object than ourselves: "the incompleteness of every singular perception is therefore complemented by the possibility for another appearance, from another point of view, *which may also be that of another subject.*" (Blomberg & Zlatev, 2014, p. 404).

Since Blomberg and Zlatev (2014) appear to assume that our ability for self-motion overlaps with transcendental intersubjectivity, they take enactive perception to favor a first-

[5] Note that affordance is always relative to the agent. A fence may afford motion for a squirrel but not stereotypically for a human being, a string may afford motion for ants, but not for people (Matsumoto, 1996), and a path affords motion for a person, but not for a bird (Blomberg, 2014).

person viewpoint. However, transcendental intersubjectivity may also be an argument against favoring any specific viewpoint, since we understand that objects may be perceived by others from *different vantage points*. Therefore, affordance of motion need not be tightly connected to our subjective first-person viewpoint, since it may just as well be the case that in my perception of a door-knob, I perceive it, not only as an object for me to interact with, but for others to interact with as well, meaning that the perception of affordance need not be linked to an enactment of myself as agent. Thus, *affordance of motion* can be distinguished from *enactive self-motion*, and both can potentially serve as distinct motivation for NAM-expressions.

2.3.1.4 Summary

I have here presented the proposed motivations for NAM-expressions, and distinguished in the third case between two separate motivations, namely *affordance of motion* which is determined by the type of object which is perceived, and *enactive self-motion*, which furthermore implies experiential enactment of the self as an agent. It remains an empirical question to determine the possible role of each motivation in the expression of non-actual motion. In the following chapters, I present an empirical study which explores the role of visual scanning, affordance of motion, and enactive self-motion as experiential motivations for NAM-expressions.

Throughout this chapter, multiple examples of NAM-expressions with traversable objects as grammatical subjects have been given. Some of these descriptions (2.6 and 2.7) were rich in manner-of-motion information, and it was suggested that they are motivated by imagination. For other NAM-expressions (2.4 and 2.8), enactive self-motion, affordance of motion and visual scanning are all possible experiential motivations. However, some NAM-expressions, with non-traversable objects as grammatical subjects, as in (2.9), may be motivated only by visual scanning.

(2.9) The fence goes along a meadow.

(Blomberg, 2014, p. 180)

In sum, the relationship between expression and motivation is an empirical question, as not all NAM-expressions may share a single motivation and a single NAM-expression may have

more than one source of motivation, calling for studies such as that in the following chapters. But first, I present a review of previous research of expression of non-actual motion in language.

2.3.2 Non-actual motion in expression

As stated at the onset of Section 2.3, the linguistic expression of NAM has been studied for a wide range of typologically diverse languages (Amagawa, 1997; Blomberg, 2014; Matsumoto, 1996; Rojo & Valenzuela, 2003, 2009; Stosic & Sarda, 2009). In each case, however, only to a limited extent, and with varying methods. The first, perhaps trivial, observation, then, is that NAM-expressions are possible in a wide range of languages, and thus constitute a possible language universal (*sensu* Coseriu, 1977). There is also some documented cross-linguistic variation, which I here review with a focus on what this evidence may tell us about the influence of linguistic conventions.

Matsumoto (1996) discusses two kinds of linguistic evidence for the experiential motivation for NAM-expressions. First, the (inherent) directionality illustrated by two "classic" examples, (2.10) and (2.11), may account for the semantic difference between the two sentences, which relies on the experience of some sort of motion in a given direction (Matsumoto, 1996).

(2.10) The mountain range goes from Mexico to Canada.

(Talmy, 2000a, p. 104)

(2.11) The mountain range goes from Canada to Mexico.

(Talmy, 2000a, p. 104)

A linguistic test has been proposed to illustrate the existence of an experiential motivation for NAM. When adding a temporal adverbial to a NAM-expression, such as *for a while*, as in (2.12), it is interpreted as describing the duration of some process, rather than the duration of existence of the grammatical subject (Matsumoto, 1996).

(2.12) The highway runs along the coast *for a while.*

(Matsumoto, 1996, p. 186, emphasis added)

While Matsumoto (1996) points out that NAM-expressions are possible in both English and Japanese, he also documents a difference with regards to the kinds of objects to which NAM-expression may be ascribed, as well as differences with regards to the types of verbs that may be used for different types of objects. According to his analysis, English may use NAM-expressions to describe "borders, wires, cables, walls, and fences" (Matsumoto, 1996, p. 214), while Japanese allows it for the first three, but not the last two. Similarly, Matsumoto found that the verbs used for NAM-expressions in English may be applied both to objects which do, and do not, afford human motion,[6] while many verbs in Japanese may only be used in NAM-expressions describing objects which do afford human motion. Such systematic differentiation has not been reported to exist in the remaining languages studied, but Blomberg (2014) did find that certain individual verbs were used consistently with only a limited range of objects.[7] Together, this suggests that certain linguistic constraints and conventions, affect the types of verbs and objects that may be used in NAM-expressions.

Blomberg's (2014) study is the most extensive study of the expression of NAM, comparing data from three different languages, Thai, French, and Swedish. The analysis is based on linguistic descriptions of drawings of spatially extended objects, which were systematically varied with respect to perspective (first-person and third-person) and affordance (traversable and non-traversable). NAM-expressions were operationally defined as "a sentence that by substituting the Figure-expression with one that denotes an object that is movable, actual motion would be the only possible interpretation" (Blomberg, 2014, p. 179).

---

[6] It should be noted, however, that there is a difference between what a language "allows", and what the speakers of a language "do". As Matsumoto's (1996) analysis employs (native) speaker intuition, it may say something about the first, but not about the second. It should therefore be noted that a language may allow a verb to be applied to a wide range of objects, while in usage it may be applied mainly to a limited range (Zlatev & Blomberg, 2019)

[7] For example, it was found that the Swedish verb *leda* ('lead') and French verb *mener* ('lead') was overwhelmingly used to describe objects which afford motion.

The overarching results of the study showed that all three languages have the possibility to express NAM, and their speakers frequently make use of it, using the same language specific resources that are used in expressions of actual motion. Interestingly, the speakers of French, Swedish and Thai produced most NAM-expressions when describing stimuli depicting traversable objects from a first-person perspective. This was taken to support the relevance of enactive self-motion as a motivation for NAM. However, given that all four conditions elicited NAM-expressions, enactive self-motion was shown not to be sufficient to motivate all occurrences of NAM-expressions.

The study also found that the motion verbs used most frequently in NAM-expressions were *generic*, such as Swedish *gå* ('go') and *leda* ('lead'), French *mener* ('lead'), and Thai *pai* ('go'). This may suggest that the motion semantics of NAM-expressions is rather bleached, as typical in cases of grammaticalization (Heine & Kuteva 2002).

However, it is interesting to note that two studies of translations of NAM-expressions between Serbian and French (Stosic & Sarda, 2009), and Spanish and English (Rojo & Valenzuela, 2003), found high degrees of retention of the NAM-meaning across translations, also in cases where alternative non-NAM translations were possible. This may suggest that the translators are attuned to the motion-semantic meaning of the NAM-expressions.

A general topic of interest in motion semantics is the expression of Manner and Path/Direction in motion-expressions. This interest derives from Talmy's (2000b) motion typology, which distinguishes between so-called "verb-framed" languages, in which Path is expressed in the verb, and "satellite-framed" languages, in which Path is expressed in a satellite to the verb. This dichotomous typology has been found to be faulted in many ways (e.g. Naidu et al., 2018; Slobin, 2004; Zlatev & Yangklang, 2004) but the notions of Path, as well as Manner (of motion) have proved to be productive, also for the analysis of non-actual motion. For example, Matsumoto (1996) proposed that the following conditions restricts the expression of Path and Manner information in NAM:

a. The path condition: Some property of the path of motion must be expressed.

b. The manner condition: No property of the manner of motion can be expressed unless it is used to represent some correlated property of the path. (Matsumoto, 1996, p. 194).

However, it is unclear what it to means for a Manner-verb to "correlate to path-features". Matsumoto gives the example of (2.13), but it is unclear how these Manner-verbs express a Manner-meaning correlating to the property of the Path, especially considering the relative wide differences in the types of manner all ascribed to the same type of path in his example.

(2.13) The path {rambles/roams/wanders} through the forest.

(Matsumoto, 1996, p. 196).

Blomberg (2014) formulated an alternative condition, which states that "the Path- and Direction-information given about a static object "overrides" the Manner-information provided by the verb" (Blomberg, 2014, p. 183).[8] This would explain why NAM-expressions that are rich in Manner-information may be interpreted as metaphors, since the semantic content of manner-of-motion is so "strong" that it may not be demoted, and thus the only possible interpretation becomes a personification of the inanimate grammatical subject (Blomberg, 2014). In (2.14) such demotion is possible, and no appeal to metaphor is necessary, while in (2.15), it may be argued that this is not so, and the only possibility thus becomes to reject the sentence or understand as a metaphor.

(2.14) A fence that runs across the field.

(2.15) A fence that tip-toes across the field.

These conditions would suggest that Path information has precedence over Manner information in NAM-expressions. Such a precedence was also reflected in a translation study, which found that Manner-information was dropped in translation, more often than Path-information (Stosic & Sarda, 2009). Relatedly, Blomberg (2014) observed an interesting tendency for Swedish participants to use expressions, which he named "Non-actual Path", in

[8] It remains an open question to what degree the manner-meaning is retained. In the case of the Swedish verb *gå* ('go'), this may be none or little, but Blomberg (2014) discusses some interesting examples such as the systematic variation between *doen* ('walk') and *wîng* ('run') in Thai with different paths on which human travel varies in speed, e.g. small roads and highways.

which no verb was included, but where Path/Direction was expressed through adverbs, as in (2.16):

(2.16)	Ett	rör	ut	genom	en	tunnel
	DET.INDF	pipe	out	through	DET.INDF	tunnel

'A pipe out through a tunnel'

(Blomberg, 2014, p. 203)

Such expressions are not really NAM-expressions since they lack motion verbs, but they may nevertheless be related to the phenomenon of non-actual motion (Blomberg, 2014, p. 189). Due to their use in expressing vantage points, as in (2.17), they appear to be motivated by visual perception (Blomberg, 2014).

(2.17)	Ett	fönster	ut	mot	havet
	DET.INDF	window	out	towards	sea-DET.DEF

'A window facing towards the sea'

(Blomberg, 2014, p. 188)

In sum, this section has reviewed a number of relevant findings. First, there is cross-linguistic evidence for the possibility to use NAM-expressions for objects that do not afford motion, which can apparently only be explained with appeal to the visual scanning motivation. At the same time, the findings discussed in this section show that this motivation cannot explain all NAM-expressions. Second, some languages place normative constraints on the types on NAM-expressions that may be used to describe traversable and non-traversable objects respectively (Blomberg, 2014; Matsumoto, 1996). Third, the cross-linguistic findings that speakers produce more NAM-expressions when describing objects that afford human motion, depicted from a first-person viewpoint (Blomberg, 2014), suggest a strong influence of the enactive-self motion motivation.

2.4 Summary and hypotheses

In this chapter, I have presented ideas such as methodological triangulation and the conceptual-empirical loop from the discipline of cognitive semiotics as the overarching theoretical and conceptual framework for this thesis. NAM-theory (Blomberg, 2014, 2015, 2017; Blomberg &

Zlatev, 2014) was introduced as an alternative to previous theories on the phenomenon of using motion expressions to describe static scenes (e.g. Talmy, 2000a; Langacker, 1987). A benefit of the NAM-theory is that it explicitly distinguishes between non-actual motion *expressions*, and non-actual motion *motivations*: visual scanning, affordance of motion, enactive self-motion, and imagination, which are different types of experiences that may motivate such expressions.

In addition, the semiotic system of *gesture* was introduced, and it was described how gestures may be analysed and categorized semantically so as to study *how* they represent the intentional objects that they are about. Thus, gestural analysis may be used to explore the influence of different possible NAM-motivations on expression, in addition to the type of linguistic analysis that has been used in the previous literature, as reviewed in Section 2.3.2. The latter suggests that the motivation for NAM-expressions is at least two-fold: visual scanning and affordance of motion. Furthermore, NAM-theory emphasises that several NAM-motivations may be co-active, and this would help explain the higher frequency of NAM-expressions in descriptions of traversable objects depicted from a first-person viewpoint (Blomberg, 2014), because these expressions may simultaneously be motivated by visual scanning, affordance of motion, and enactive self-motion. In accordance with this, it may be the case that the more possible motivations exist for any given stimulus, the higher the chance that the speaker experiences any of them, which might motivate the use of a NAM-expression. By considering only linguistic evidence, it seems impossible to tease apart these possibilities.

Based on the theoretical background, and findings of previous empirical studies, this thesis proposes the following two general hypotheses:

- **H1** The more NAM-motivations that may apply to a given static scene, the more likely it is to be described with a NAM-expression in speech and gesture.
- **H2** Different NAM-experiences may be reflected by different types of iconic gestures.

In addition, an exploratory question deserves attention, given that this is this first time NAM-expressions are studied through two different semiotic systems:

- Do NAM-expressing gestures tend to co-occur with linguistic NAM-expressions, or with linguistic descriptions without NAM-expressions?

3 Methodology

In Chapter 2, I argued that there are multiple possible NAM-motivations, and that NAM-expressions are an integral part of many, and possibly all human languages. However, the exact nature of the relationship between experience and expression of non-actual motion remains an open question. To address this, the present chapter introduces a study of linguistic and gestural expressions of non-actual motion in Swedish, which is based on the experimental design of a previous cross-linguistic study of NAM (Blomberg, 2014), and thus turns the focus to the empirical side of the conceptual-empirical loop (see Section 2.1.1)

The stimuli consisted of drawings that were systematically designed to favor different NAM-motivations, which were used to elicit descriptions in both speech and gesture. The data that I have analysed is part of a larger dataset consisting of data from four languages: Thai, French, Bulgarian and Swedish, and four different age-groups: 5-, 7-, 9- year olds, and adults, which was collected as part of the project *Phenomenology and Typology of Motion* (PATOM) (Zlatev, Blomberg, Devylder, Naidu & van der Weijer, submitted).[9]

3.1 Participants

Thirty-nine native Swedish speakers participated in the experiment (20 adults and 19 nine-year old children).[10] The data from two participants (1 adult, 1 child) was excluded due to lack of

[9] See the website of the project for more information: https://projekt.ht.lu.se/en/patom/

[10] Two age-groups were included in this study, but no developmental hypotheses was tested. Data from the 9-year olds was included to increase the total amount of data. Previous research on gesture acquisition show increased gestural skills based on a range of parameters in children at this age, compared to younger children (Capirci et al., 2011), and it has been shown that gestures of older children (6-year olds, compared to 4-year olds) are more comparable to adult gesture production (Gullberg, Hendriks & Hickmann, 2008).

descriptions for parts of the stimuli-set. The data of 37 participants was therefore included in the analysis, 19 adults (12 female, ean-age 29.5), and 18 children (10 female, mean-age 9.0). The children were recorded at their school; their participation was voluntary, and with the permission of their legal guardians, who provided written informed consent. The adult participants were recruited with the help of posters that were placed at various places in Lund, and they were recorded at Lund University. All signed an informed consent form and were compensated with a cinema ticket.

Due to the time-consuming nature of gesture annotation, only a subset of the videos were analysed for gesture. The subset consisted of half the participants; ten adults (5 female, mean age 25.3), and nine children (4 female, mean age 9.0).

3.2 Materials

Forty color drawings (20 target, 20 control) were used as elicitation material in this study (see Appendix A). The 20 target drawings depicted ten elongated, static objects in natural contexts, such as a road through a cave, or a fence across a field (see Figure 11). Half the target-objects were meant to afford motion, while the other half was not. Each object was depicted twice, once from a first-person perspective,[11] and once from a third-person perspective. The 20 control drawings depicted everyday objects with none or little elongation, also in natural contexts, such as an apple on a tree or a volcano surrounded by clouds (see Figure 12).

[11] Following Blomberg (2014, p. 175), some [1pp, -afford] stimuli were depicted at a slight angle, to avoid possible interference in their descriptions caused by the unnaturalness of a 90-degree angle, giving the experience of standing or, as one participant in the present data described it, for the 1pp perspective of the chain, “lying on top” of the object.

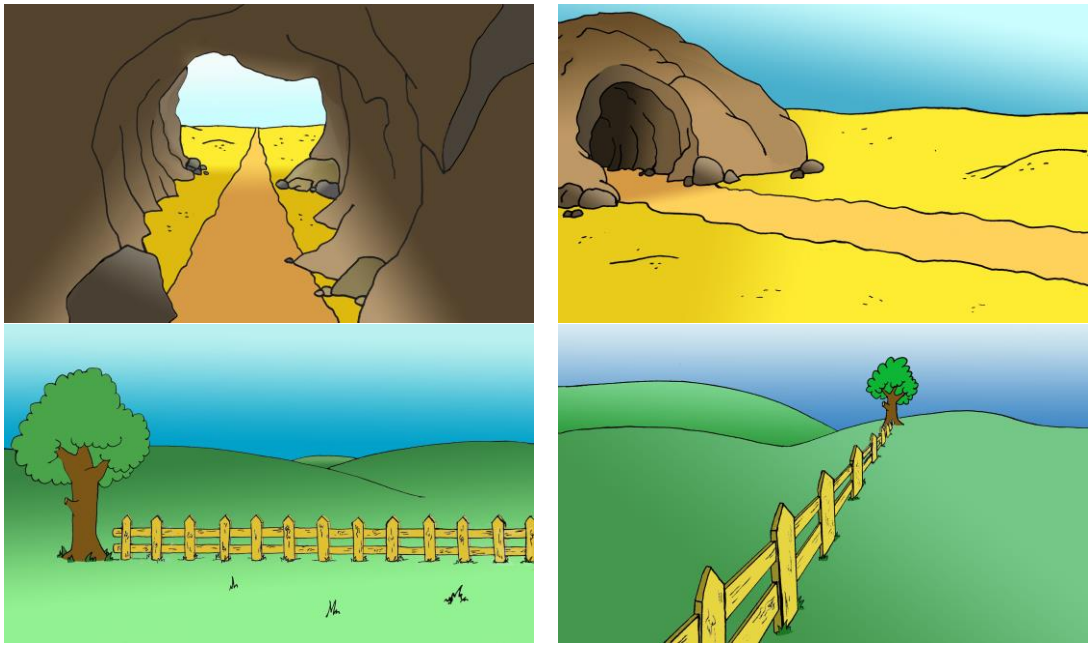


Figure 11 - Examples of target stimuli.
Top (from left): 1pp afford, 3pp afford. Bottom: 1pp non-afford, 3pp non-afford

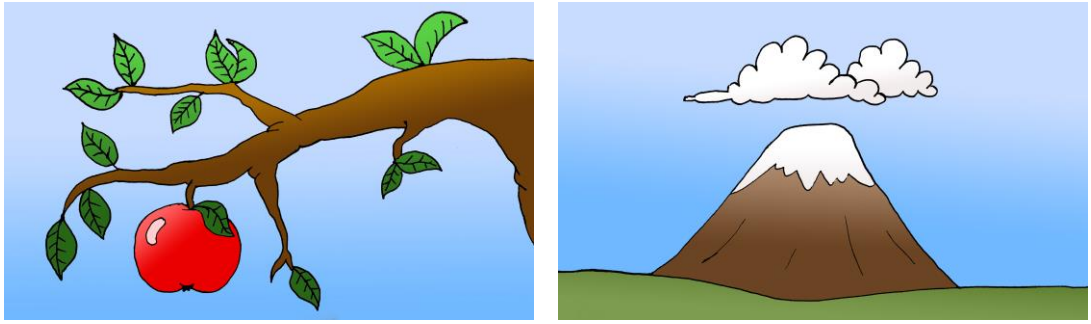


Figure 12 - Example of control stimuli

3.3 Elicitation procedure

Elicitation took place in a quiet room, and was video-recorded. Participants sat in a chair without armrests in front of a computer monitor controlled by the experimenter to show stimuli-drawings one at the time. The stimuli were randomized and presented using the E-prime software (Psychology Software Tools, Pittsburgh, PA).

An assistant to the experimenter was seated across from the participant, and a camera was positioned to capture both the participant and the assistant from the side, as illustrated in Figure 2. The experiment was presented as a game to be played by the participant and assistant. Participants were asked to describe the drawings for the assistant to identify, and mark, on a sheet containing miniatures of all stimuli-drawings. Participants were instructed to give short, but clear descriptions of the pictures, which would allow the assistant to identify it. The sheet was briefly shown to the participants before the task started.

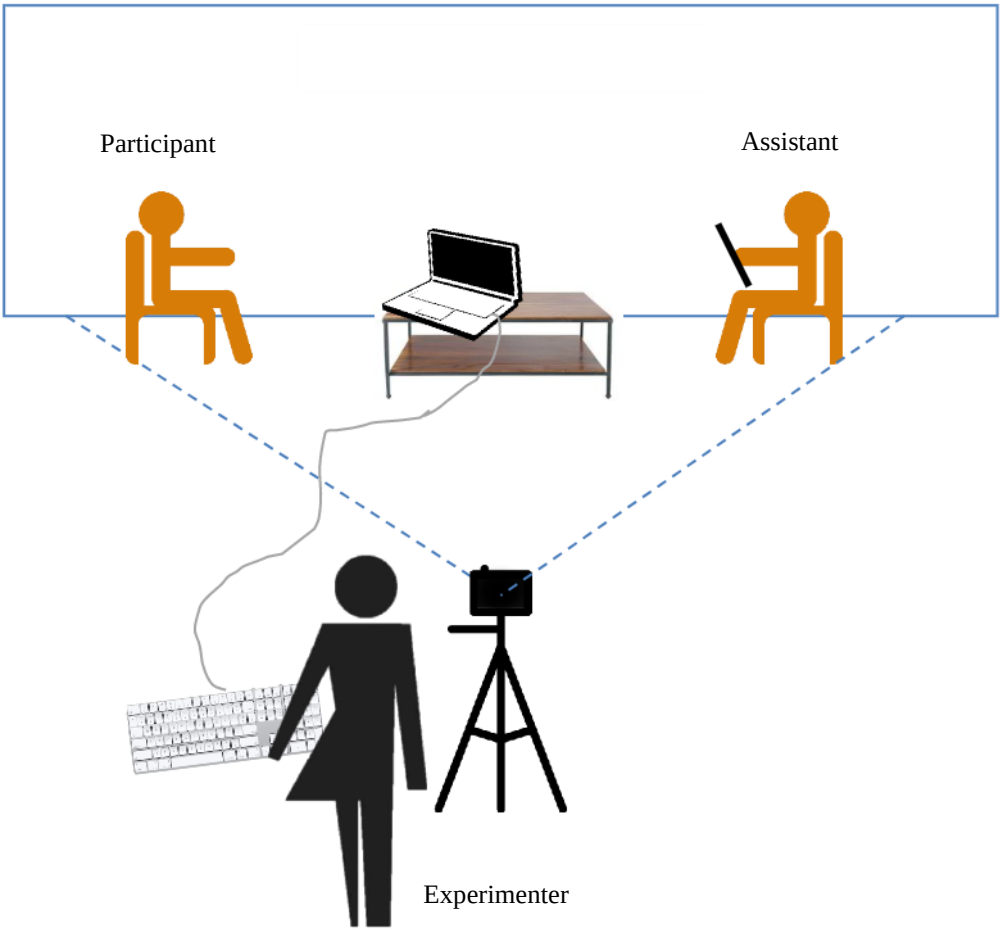


Figure 13 - Experimental set-up

Prior to starting the task, the participants were given five training-trials. The experimenter gave the participant feedback on the length of their descriptions during training, and after

completing the five training-trials the assistant showed them their marked sheet, to highlight the nature of the game. During elicitation, participants did not receive feedback, beyond a signal from the assistant that they had located the drawing. If participants gave one-word descriptions, the assistant would not immediately mark a picture, to encourage them to say more, and if necessary the experimenter would ask the participant to describe the picture a bit more.

3.4 Speech coding

The speech was transcribed in ELAN (Wittenburg, Brugman, Russel, Klassmann, & Sloetjes, 2006) and segmented into clauses containing a single finite verb by a native Swedish speaker. A custom "lexicon" was produced, consisting of all lexemes extracted from the transcriptions. Each lexeme was marked for form-class, of which verbs are of relevance to the present thesis. Entries were further categorized according to the theory of *Holistic Spatial Semantics* (Zlatev, 1997), which proposes distinct semantic categories that encode spatial meaning and motion semantics (Blomberg, 2014; Naidu et al., 2018). Of relevance here, is the category Motion, which is a binary category marking the presence or absence of perceived motion, defined as continuous change of position against a background (Zlatev et al., 2010, see Chapter 1). This lexicon allowed the automatic identification of all motion verbs in the data.[12]

Given the broad definition of NAM-expressions: *an expression using a motion verb to describe a static scenario* (Blomberg, 2014), all clauses in the dataset containing a motion verb may be taken to be NAM-expressions, seeing as they were all produced as descriptions of the static scenarios depicted in the stimuli. However, to ensure comparability to previous research with a similar methodology, which employed a more narrow operationalized definition of NAM-

[12] With the application of such lexicon, we may consider this part of the analysis to be performed from a third-person perspective, as it is independent of the author's insights and intuitions. However, the production of the lexicon itself, relies on first-person intuitions and intersubjective second-person confirmations that these intuitions are socially shared.

expressions (Blomberg, 2014; Blomberg & Zlatev, 2015), NAM-expressions were further subcategorized into *narrow* and *extended* NAM-expressions. Following Blomberg (2014) and Blomberg and Zlatev (2015), cases where replacing an inanimate, static entity as the subject of a motion-verbs with an animate, self-moving entity led to the expression of actual motion, as exemplified in (3.1), were coded as narrow NAM-expressions (N-NAM-E). In some cases, sentences were not considered acceptable when replaced with a single entity, but replacement with a mass of autonomous, moving entities provided acceptable sentences, as exemplified in (3.2). Such sentences were also included as narrow NAM-expressions.

(3.1) [*The road/Stanislaw Lem*] goes into a tunnel.

(Blomberg, 2014, p. 180, emphasis added)

(3.2)

[*solljuset/myrorna*]	strömmar	in	mot	mig.
[the sunlight/the ants]	stream(s)	in	towards	me[13]

'The [sunlight is/ants are] streaming in towards me.'

(D74_Light1pp)[14]

Extending on this narrow definition, the remaining cases were collected under the term extended NAM-expressions (X-NAM-Es)[15].

3.5 Gesture coding

Gesture coding was performed by the author, who identified all occurrences of gestures in the video-recorded data[16], and marked them in ELAN. A gesture was identified as a bodily action

[13] A simplified gloss is included with Swedish examples from the data.

[14] All examples from the data will be marked with a code referencing the participant ID (where C= 9-year olds, and D = adults), and the stimuli name (where all target stimuli will be marked as either 1pp or 3pp). Due to space constraints, only the relevant part of the stimulus description is included.

[15] The coding of NAM-expressions into narrow and broad categories was based on first-person intuitions of the acceptability of the replaced sentences. It should be noted that I am not a native speaker of Swedish, but in cases of doubt, I have conferred with native Swedish speakers.

that was found to be either explicitly communicative, representationally complex to constitute a sign, or both, following the definition given in Section 2.2.1.[17]

Subsequently, all gestures were coded, binarily, as iconic or not. As discussed in Sections 2.1.2 and 2.2.1, gestures, and signs in general, may have more than one semiotic ground, but gestures that were found to have a predominantly iconic ground were coded as iconic from a first-person perspective, based on systematic intuition.[18] Non-iconic gestures were not coded for their specific semiotic ground, as this was not relevant for the current study. While indexical and emblematic gestures where not coded as such, all three possible grounds were considered contrastively when making ground-judgements.

All gesture coding was performed with the sound enabled and I had access to the relevant stimulus drawing, which was employed to assist in determining the semiotic ground when necessary. This was relevant, as this study was not interested solely in gestures displaying traits of primary iconicity, but also gestures of secondary iconicity, as discussed in Section 2.2.1, which required additional means to establish the ground relation between the gesture and the intentional object. In these cases, the full communicative context, including speech and pictures were used to establish this relation in coding.

---

[16] For practical reasons, and due to time constraints, gestures performed with the head as articulator were not coded, as many of these consisted of head-nods and -shakes, which were not relevant to the topic of this thesis.

[17] This coding was performed from a second-person, empathetic, perspective (cf. Section 2.1.1), where I considered the gestures produced by participants within the communicative context and as means for intersubjective communication. Due to time-constraints, secondary coding by another researcher was not possible, but would have been beneficial to establish that coding was indeed intersubjectively consistent.

[18] While the identification of gestures relied on an empathetic, second-person perspective, based on the general idea that communicative intent may be experienced by others (for a discussion of second-order communicative intent and intimations, see Section 2.2.1), it is less straight-forward to determine the “actual” meaning of a gesture. This coding was therefore performed from a first-person perspective applying systematic intuitions, in determining the presence or absence of similarity between expression and (assumed) intentional object. The introduction of a secondary coder (see footnote 17), would have allowed the comparison of intuitions, introducing a second-person, intersubjective perspective.

All iconic gestures were categorized based on a coding scheme, developed in conjunction with the PATOM project, and adapted for the purpose of this study. This was intended to distinguish between gestures that do and do not express motion, and further, between different kinds of motion expressions (see Table 4). This coding scheme was developed on the basis of the conceptual-empirical loop (see Section 2.1.1), where the categories were not pre-determined prior to working with the gesture-data. It was developed considering both "top-down" semantic categorization of gestures and "bottom-up" observations of the actual data.

The scheme consists of five iconic gesture types, which may be taken to express motion to increasing degrees:

1. Shape/placement (SP): Gestures representing the location or shape of something. This category includes all gestures with static nuclei. During preparation the hand(s) are brought to the apex, where they are held in a way that is semantically meaningful in the context.

2. Extension (E): Gestures representing the shape of some static entity, by dynamic means, for example by outlining it.

3. Trace (T): Gestures representing the directionality of something or a path of motion.

4. Motion (M): Gestures representing the actual motion of some entity.

5. Motion-manner (MM): Gestures representing some manner of motion, for example walking, jumping, etc. in general, or of some entity.

The classification of iconic gestures into these five types was based of three formal and semantic features: (a) the stroke type, (b) the number of vectors, and (c) the type of hand-mapping. As discussed in Section 2.2.3, a stroke may be either static (also called stroke-less hold, Kita et al., 1999) or dynamic, where static strokes are the least likely candidates to express motion. However, even dynamic strokes may be used to represent static objects for example in the case of "outlining" gestures (Brown, Mittermaier, Kher, & Arnold, 2019), or in what Müller (2014) called "drawing" gestures (see Section 2.2.2), which indicates that not all dynamic strokes may be taken to express motion.

Therefore, a further feature employed to distinguish between gesture-types was the number of vectors expressed in the gesture. A *vector* was defined as a directional movement along an axis, which means that it includes not only linear movements as single vectors, but also complex movements along one axis, such as zig-zagging or spiralling (see Figure 14). Given the typical goal-directed nature of human motion, if was considered that single-vector gestures would be more likely to represent motion, whereas multi-vectorial gestures may rather be used to indicate the outline of something. This conjecture is of course tentative, and relative to the current stimuli, which includes straight, simple paths. A gesture representing a complex path, e.g. “across the river, over the hill, and to the right” might well be multi-vectorial, but was considered unlikely to occur in the present dataset.

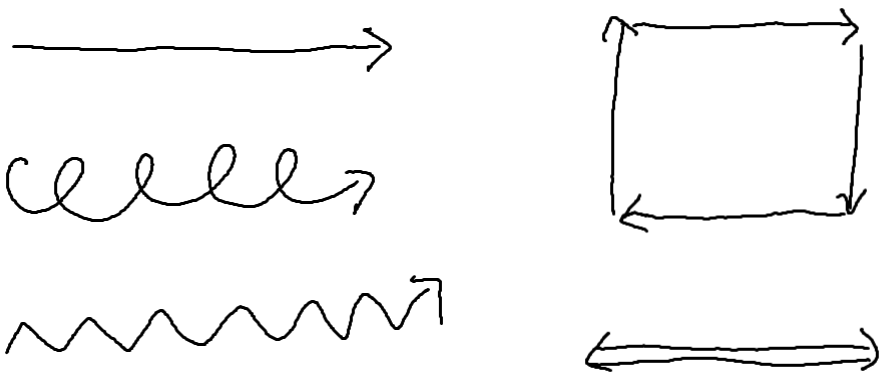


Figure 14 - Illustration of single vectors (left) and multi-vectors (right)

Lastly, the hand-mapping types proposed by Cartmill et al. (2017), introduced in Section 2.2.2, were used to help identify gestures that express explicit agents of motion. For gestures produced with hand-as-hand, the speaker would be likely to use their body to express the motion of themselves or an embodied other. For dynamic gestures produced with hand-as-object, the movements of that hand can be expected to represent the movement of the represented object. Finally, in hand-as-neutral gestures, the agent is underspecified, meaning that these gestures provide a less explicit expression of motion. The summary of these considerations, and the basis for the coding scheme is given in Table 4.

Table 4 - Motion-categories for gesture coding, along with criteria for distinguishing them

Gesture category	Stroke-type	Vectors	Hand-mapping
Shape/Placement (SP)	Static	None	Hand-as-neutral, hand-as-hand, or hand-as-object.
Extension (E)	Dynamic	Two or more	Hand-as-neutral
Trace (T)	Dynamic	One	Hand-as-neutral
Motion (M)	Dynamic	One or more linear vector(s)	Hand-as-object, or hand-as-hand
Motion-manner (MM)	Dynamic	One or more complex vector(s)	Hand as object, or hand-as-hand

To exemplify, the representation of a ladder by presenting the arm in a diagonal line in front of the torso (as illustrated in Figure 6), would be coded as a Shape/Placement gesture (static stroke, no vectors, hand-as-object). The upward-going index-finger tracing (Figure 8), would be coded as a Trace gesture (dynamic stroke, one vector, hand-as-neutral). Had this stroke consisted of repeated movements up- and downwards, it would have been coded as an Extension gesture (dynamic stroke, two or more vectors, hand-as-neutral). Both the gesture representing movement up a ladder with the index-fingers as legs (Figure 7) (dynamic stroke, one complex vector, hand-as-object), and with the hands grabbing the rungs (Figure 5) (dynamic stroke, one complex vector, hand-as-hand), would be coded as Motion-manner gestures, whereas a gesture performed with a handshape of index-finger and middle-finger extended in a downwards V-shape to represent legs, where the whole hand is moved upwards in a straight line, would be coded as a Motion gesture (dynamic stroke, one linear vector, hand-as-object).

As Trace, Motion and Motion-manner gestures were taken to be expressive of motion, they may be considered as expressive of non-actual motion when employed in descriptions of the static scenes presented in the stimulus of this study. Thus, in the following I refer to these,

collectively as NAM-expressing gestures, or for short: *NAM-gestures*, and the remaining two, Shape/Placement and Extension as *static gestures.*

3.6 Predictions

Based on the general hypotheses presented in Section 2.4, and the methodological refinements presented in the present chapter, we may now formulate the following specific predictions. To remind, Hypothesis 1 stated that the more possible NAM-motivations may apply to a scene, the more likely it is that it will be described using a NAM-expression. On this basis, we can formulate the following prediction:

- **P1** The frequency of NAM-Es will be as predicted in Table 5, where the stimuli type in the first row [1pp, +afford] is predicted to elicit the greatest amount of NAM-expressions (in both speech and gesture), and the amount of NAM-expressions will decrease for each subsequent row.

Table 5 - A graphic formulation of Prediction 1 (see the main text)

Stimuli	Possible NAM-Motivations
[1pp, +afford]	Enactive self-motion, Affordance of motion, Visual scanning
[3pp, +afford]	Affordance of motion, Visual scanning
[1pp, -afford] **[3pp, -afford]**	Visual scanning
[control]	None

Highest frequency of NAM-Es

Lowest frequency of NAM-Es

Section 3.5 introduced the scheme for classifying iconic gestures, based on their different ways of expressing motion. To remind, Hypothesis 2 stated that different NAM experiences will be reflected in different types of gesture. We may now further elaborate this hypothesis and state that Motion-Manner gestures would reflect the enactive self-motion motivation, Motion gestures

would reflect the affordance of motion motivation, and Trace gestures – the visual scanning motivation. This can be formulated as the following predictions:

- **P2a** Motion-manner (MM) gestures will be produced for [1pp, +afford] stimuli;
- **P2b** Motion (M) gestures will be produced for [+afford] stimuli;
- **P2c** Trace (T) gestures will be produced for all target stimuli.

In addition to these predictions, the interaction between speech and gesture in expression of non-actual motion is explored. It may be the case that speakers enlist either speech or gestures, individually, to express non-actual motion. Alternatively, speech and gesture may be implored in conjunction. It is therefore asked, as an exploratory question, how frequently NAM-gestures are produced within picture descriptions employing NAM-expressions in speech.

4 Results

4.1 Speech

The dataset consisted of 3012 clauses (1149 produced by the children, 1863 by the adults), of which 410 contained at least one motion verb[19], meaning that they were classified as NAM-expressions in the broad sense. Of these, 376 were coded as narrow NAM-expressions (N-NAM-Es), and 34 were coded as extended NAM-expressions (X-NAM-E). Of the 1480 picture descriptions, 382 contained at least one NAM-expression (25.8%), and the majority of NAM-expressions were used to describe target stimuli, as shown in Figure 15.

All participants, with one exception, used at least one NAM-expression in their descriptions. In addition, five participants (all children) used less than five NAM-expressions (see Figure B1 in Appendix B). All target stimuli elicited NAM-expressions (see Figure B2 in Appendix B). Taken together, this show that NAM-expressions are not a marginal phenomenon in Swedish, and that the design of the stimuli succeeded in eliciting these expressions, with a clear distinction between target and control stimuli.

[19] Following Zlatev et al. (2010) this includes both translocative verbs (e.g. fly) and non-translocative verbs (e.g. wave).

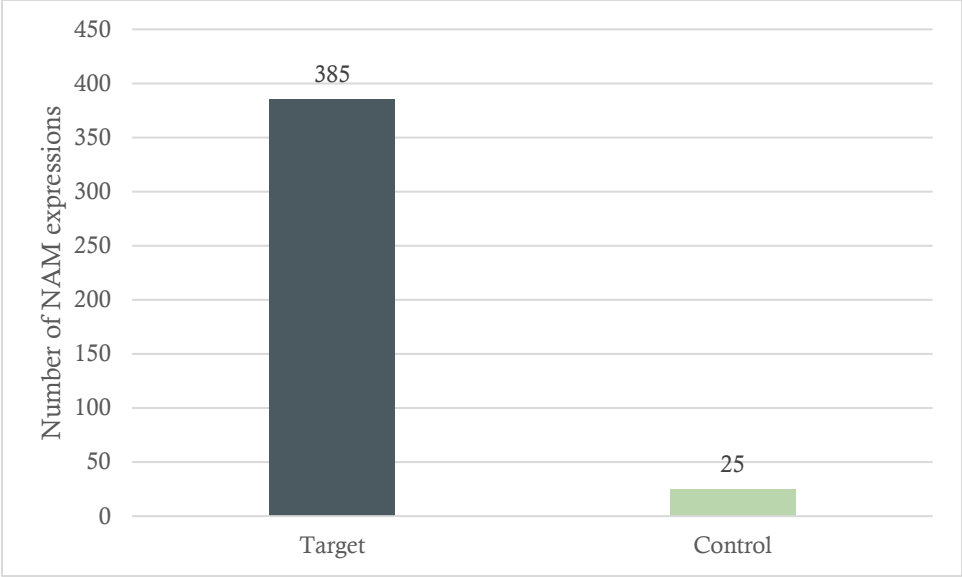


Figure 15 - Distribution of NAM-expressions across target and control stimuli.

All four types of target stimuli elicited NAM-expressions, as shown in Figure 16, with the main difference in frequency being between the affordance and non-affordance stimuli. The stimuli that invite both affordance of motion and visual scanning motivations ([1pp, + afford] and [3pp, +afford]) elicited more NAM-expressions than stimuli that invite only visual scanning ([1pp, -afford] and [3pp, -afford]). Within the types of stimuli that do not afford motion ([1pp, -afford] and [3pp, -afford]) an unpredicted difference between stimuli presented from a first-person viewpoint and a third-person viewpoint was observed.

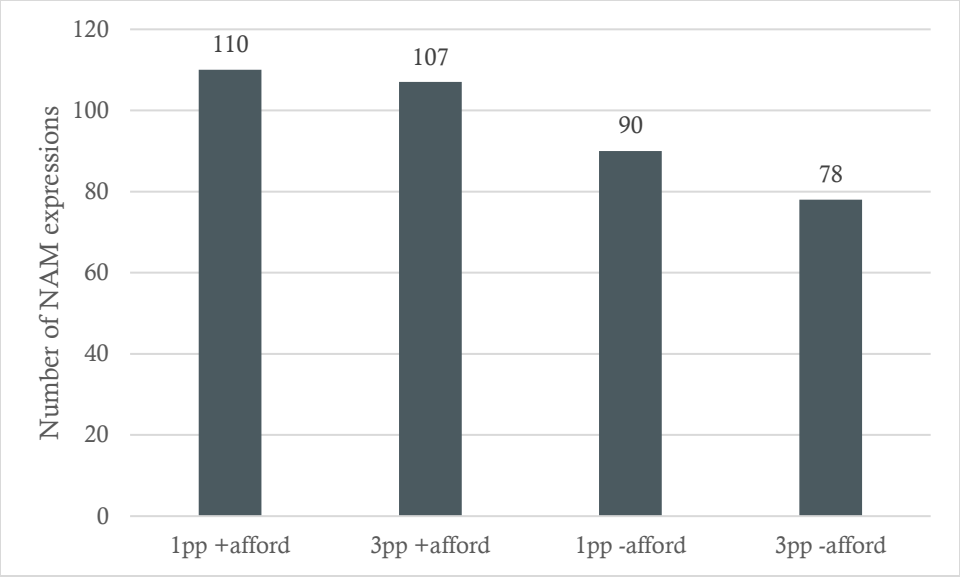


Figure 16 - Distribution of NAM-expressions across the four types of target stimuli.

4.1.1 Narrow NAM-expressions

As for NAM-expressions broadly, the majority of narrow NAM-expressions were used in descriptions of the target stimuli, as can be seen in Figure 17.

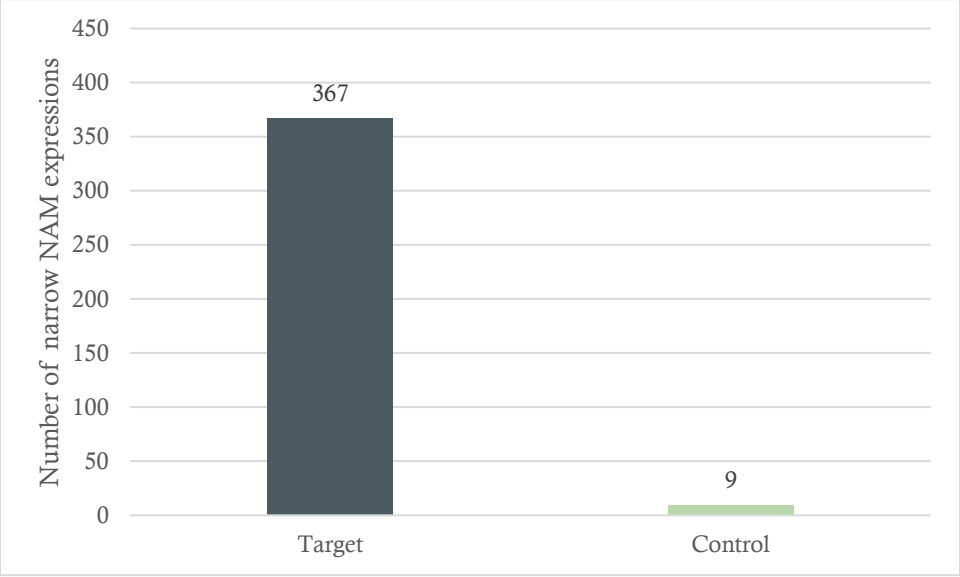


Figure 17 - Distribution of narrow NAM-expressions across target and control stimuli.

The N-NAM-Es were also elicited by all four types of target stimuli, as seen in Figure 18, with a smaller difference between those stimuli which afford, respectively do not afford motion. Within the affordance stimuli, the stimuli depicted from a third person viewpoint elicited a greater number of narrow NAM-expressions. As found for the analysis of NAM-expressions broadly, N-NAM-Es also displayed the unpredicted difference between [1pp, -afford] and [3pp, -afford].

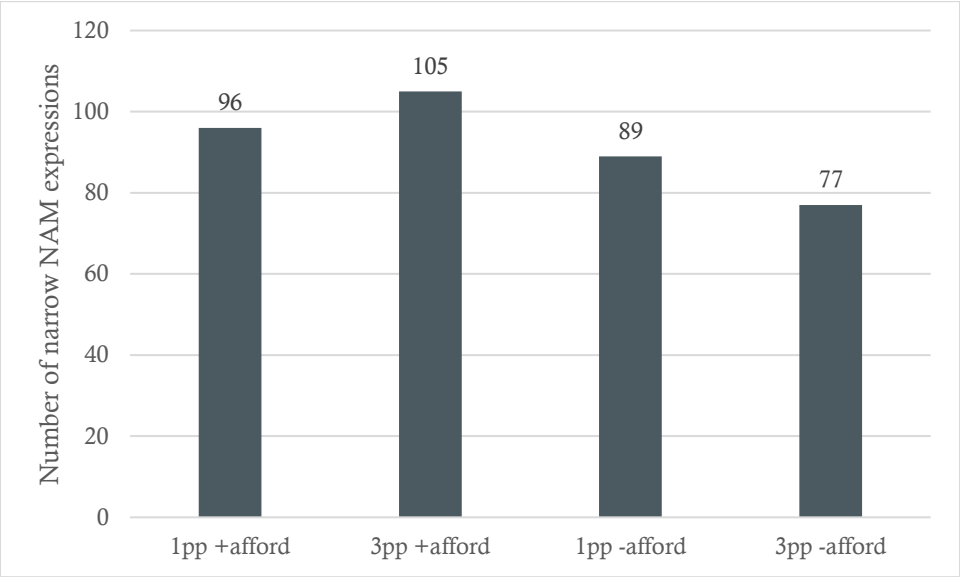


Figure 18 - Distribution of narrow NAM-expressions across the four types of target stimuli.

Little more than half of the N-NAM-Es were produced with the generic motion verb *gå* (‘go’) (53.2%), as shown in Table 6. It is interesting to note that the second most frequent verb, *leda* (‘lead’), was used nearly twice as frequently for [+afford] stimuli (61.1%), than for [-afford] stimuli, a tendency also found in a previous study (Blomberg, 2014, p. 216).

Looking at the column for [1pp, -afford] we may notice that many instances of single-use verbs are occurring here, as well as half of the total instances of the verb *komma* (‘come’) which may account for the increased usage of NAM-expressions used for [1pp, -afford] compared to [3pp, -afford]. Closer inspection exposed that the majority of usages of *komma* was used in descriptions of the 1pp stimuli where the figure was sunlight (see Appendix A), which also accounts for the single usages of *landa* (‘land’), *strömma* (‘stream’), and *tränga* (‘permeate’). It is not completely clear why the first-person depiction of the stream of light elicits more NAM-

expressions than the third person-depiction, but it may be related with the deictic meaning of *komma*, as being directed towards the speaker.

In this regard, we may notice that the sunlight stimuli differ from the remaining ones in the sense that the picture does present, in some sense, a static depiction of actual motion, namely of particles moving. However, this does not fall within the experiential definition of motion, as notions such as “photons” are not part of the perceived life world. However, this finding suggests that there is some overlap between the expression of non-actual motion, and the expressions of non-perceivable motion, which may be of interest to future research.

Table 6 - Motion verbs (frequency) used to describe each stimulus type.

[1pp, +afford]	[3pp, +afford]	[1pp, -afford]	[3pp, -afford]	[control]
Gå (‘go’) (47)	Gå (‘go’) (53)	Gå (‘go’) (48)	Gå (‘go’) (46)	Gå (‘go’) (6)
Leda (‘lead’) (42)	Leda (‘lead’) (48)	Leda (‘lead’) (18)	Leda (‘lead’) (17)	
Komma (‘come’) (1)	Komma (‘come’) (3)	Komma (‘come’) (9)	Komma (‘come’) (2)	Komma (‘come’) (3)
Sträcka (‘stretch’) (4)	Sträcka (‘stretch’) (1)	Sträcka (‘stretch’) (6)	Sträcka (‘stretch’) (6)	
		Träffa (‘meet’) (1)	Träffa (‘meet’) (3)	
		Löpa (‘run’) (1)	Löpa (‘run’) (1)	
		Kasta (‘throw’) (1)	Kasta (‘throw’) (1)	
Köra (‘drive’) (1)		Landa (‘land’) (1) Spricka (‘crack’) (1) Strömma (‘stream’) (1) Tränga (‘permeate’) (1) Åka (‘travel (by vehicle)’) (1)	Spänna (‘fasten’) (1)	

4.1.2 Extended NAM-expressions

The extended NAM-expressions occurred with similar frequency for target and control stimuli, as shown in Figure 19, and the majority of X-NAM-Es used for target stimuli describe [1pp, +afford] stimuli. Closer analysis showed that the usage of the expressions differ between the control and target stimuli. First, only five control stimuli elicited X-NAM-Es, and the majority were used to describe just two pictures, whereas nine target stimuli elicited X-NAM-Es, as seen in Figure 20.

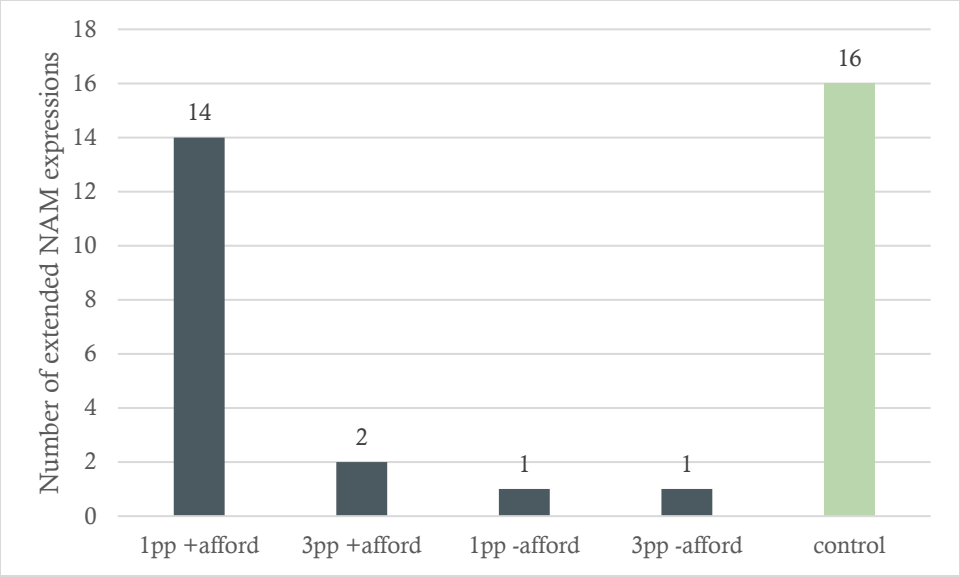


Figure 19 - Distribution of extended NAM-expressions across stimuli types.

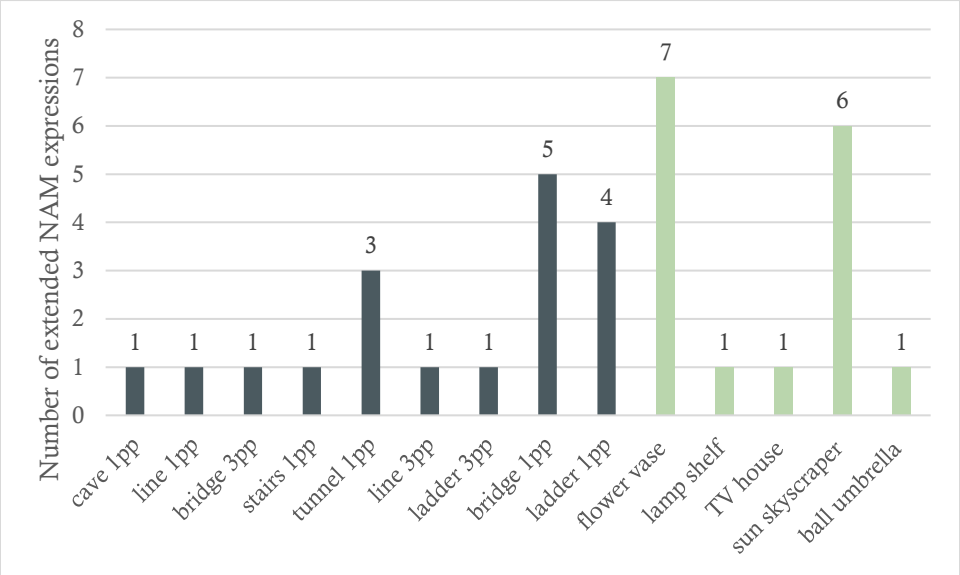


Figure 20 - Distribution of extended NAM-expressions pr. picture. Pictures eliciting no X-NAM-Es not included.

Second, the types of entities used as grammatical subject in these expressions differed between those used in description of target and control stimuli. In 17 of 18 X-NAM-Es describing target stimuli, human beings who are not depicted in the stimuli, but rather superimposed in the situation, are posed as grammatical subjects, using the pronouns *jag* ('I'), *vi* ('we'), or *man* ('one'). In 15 of 16 X-NAM-Es used to describe the control stimuli, the

grammatical subject of the motion verbs are non-human objects that are depicted in the stimuli. This indicates that the function of the X-NAM-Es used for the target- and control stimuli differ.

Two types of X-NAM-Es with human beings as grammatical subjects were used, mainly to describe the target stimuli. The first kind consisted of descriptions as if someone is moving through the scene (11 [1pp, +afford]), using either the impersonal *man* (‘one’) or the first person singular or plural pronouns *jag* (‘I’) and *vi* (‘we’), as illustrated in example (4.1), (4.2), and (4.3). This type of X-NAM-E may be called Imaginary (I-NAM).

(4.1)	En tunnel	sett	inifrån		tunneln	då	som om		
	A tunnel	seen	from the inside of		the tunnel	then	as if		
	man	skulle	gå	på	en gang	ut	från	tunneln.	
	one	should	walk	on	a path	out	from	the tunnel	

‘A tunnel seen from the inside as if one is walking on a path out of the tunnel.’

(D41_cave1pp)

(4.2)	Det	är	den där stegen	fast	nu	är	det	som	att	jag
	It	is	that ladder	but	now	is	it	as	that	I
	klättrar	upp	för	stegen	till	fönstret	på			
	climb	up	for	the ladder	to	the window	on			
	ett hus.									
	a house									

‘It is that ladder, but now it is like I’m climbing up the ladder to the window of a house.’

(C57_ladder1pp)

(4.3)	En brygga	som	vi	står	framför	som	att	vi	ska
	A bridge	that	we	stand	in front of	as	that	we	shall
	gå	igenom	det.						
	walk	through	it						

‘A bridge that we stand in front of, like we are about to walk through it.’

(D38_bridge1pp)

The second type of X-NAM-E with superimposed subjects were used to specify what an object may be used for (1 [1pp, -afford], 1 [1pp, +afford], 1 [control]), using either the construction *man kan* (litt: ‘one can’), or *man användar* (litt: ‘one uses’), as exemplified in (4.4). This type of X-NAM-E will be called Potential (P-NAM). It is perhaps not surprising that this type of expression was applied to various stimuli types, given that they all depict everyday items that people regularly interact with.

(4.4)	Och	så	ser	jag	en blå byrå	med	röda knappar	som
	And	then	see	I	a blue dresser	with	red buttons	that

man	använder	till	att	dra	ut.
one	use	to	to	pull	out

‘And then I see a blue dresser with red buttons that you use for pulling out.’

(C57_lampShelf)

The remaining four X-NAM-Es with superimposed grammatical subjects (3 [1pp, +afford], 1 [3pp, +afford]), all had different constructions which may also be interpreted to express the typical usage or function of an object.

Two types of X-NAM-Es were used with a non-human object depicted in the stimuli as their grammatical subject. These were primarily used to describe the control stimuli. One described a static situation as the result of a motion (9 [control], 1 [3pp, -afford]), either in the form of an uncaused motion of the object itself, as in (4.5), or a caused motion, without providing explicit information of the agent, as in (4.6). This type of X-NAM-E will be called Resultative (R-NAM)

(4.5)	En vas	på	ett brunt bord	där	blommorna	har	ramlat	ut.
	A vase	on	a brown table	there	the flowers	have	fallen	out

‘A vase on a brown table, where the flowers have fallen out.’

(D54_flowerVase)

(4.6)	Ett beige och rött hus	med	öppen	dörr	och	tv:n utslängd.
	A beige and red house	with	open	door	and	the TV thrown out

‘A beige and red house with the door open and the TV thrown out.’

(C56_tvHouse)

The second type described on-going motion of an object depicted in the picture, which indicate that the picture was interpreted to present a “snapshot” of a motion event (6 [control]). This type of X-NAM-E will be called Simulated (S-NAM), and is exemplified in (4.7).

(4.7)	En stad	där	solen	går	ner	bakom	ett höghus.
	A city	there	the sun	goes	down	behind	a skyscraper

‘A city where the sun is setting behind a skyscraper.’

(C46_sunSkyscraper)

4.1.3 Comparison of children and adults

The results reported in this chapter were produced by both children and adults. To ensure that differences between descriptions produced by the two groups did not affect the results, the data for the two groups was compared. Overall, the adults produced twice as many NAM-Es as the children (adult 69%, children 31% of total NAM-Es). For the N-NAM-Es, the pattern of distribution across the five types of stimuli was similar (see Figure 21), with some differences in the specific distribution, where the children produced proportionally more N-NAM-Es for the [3pp, +afford] stimuli, while the adults did so for [3pp, -afford].

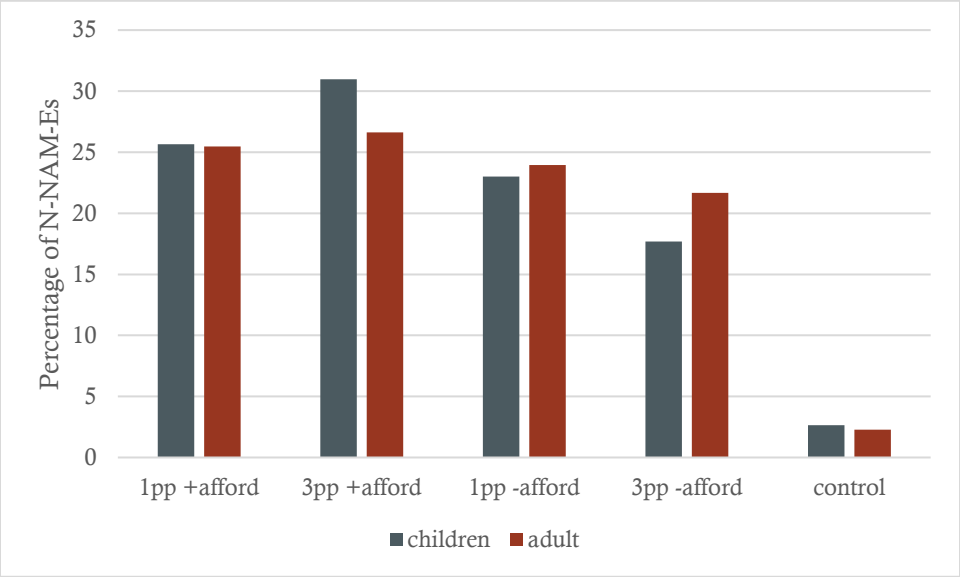


Figure 21 - Proportional distribution of narrow NAM-expressions produced by children and adults across stimuli types.

Looking at the distribution of the X-NAM-Es in Figure 22, it can be seen that children and adults produced opposite patterns, with adults producing most X-NAM-Es for [1pp, +afford] stimuli and fewer for [control], while children produced most X-NAM-Es for [control] and fewer for [1pp, +afford].

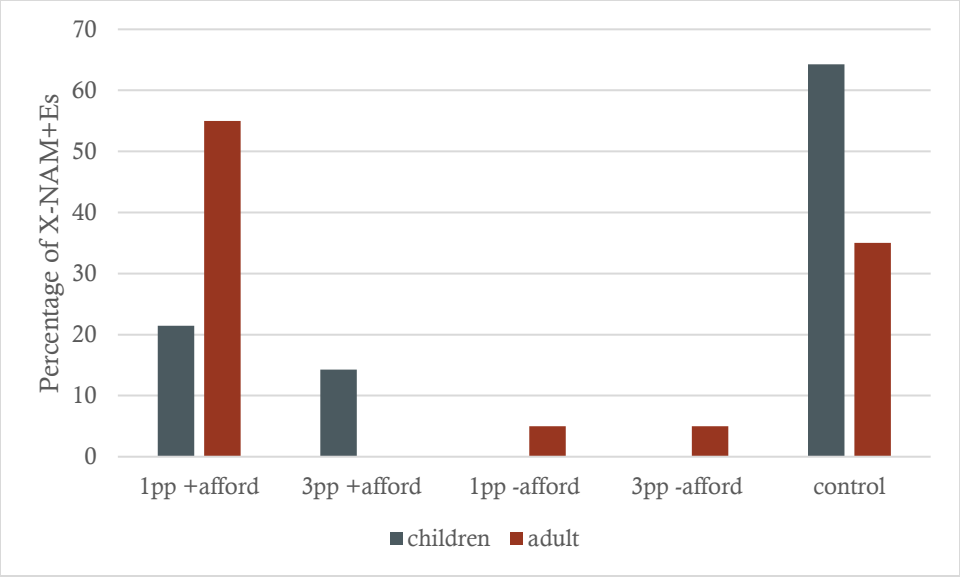


Figure 22 - Proportional distribution of extended NAM-expressions produced by children and adults.

These results will not be further discussed in Chapter 5, as no hypotheses or research question were posed to address the difference between children and adults. Here, I briefly present a possible confounding factor, and point to a potential difference between narrow and extended NAM-expressions, that should be addressed in future research. The larger proportion of NAM-Es produced by adults could be a consequence of adults producing a higher number of clauses in general in their description of the stimuli, and indeed it is the case that 15.2 % of clauses produced by adults contained a NAM-E, while that is true for 11.1 % of clauses produced by children. The finding that N-NAM-Es were distributed similarly across the four types of target stimuli for both groups, while X-NAM-Es varied, might suggest that the usage of N-NAM-Es is fairly conventionalized, and acquired before the age of nine. On the other hand, X-NAM-Es were used quite differently by children and adults, which might suggest that these expressions are more dependent on an active processes of conceptualization of the picture, including the assignment of prominence to certain aspects of the stimulus over others, where adults and children may differ, with adults favouring enactive self-motion, and children favouring the simulated perceived motion of depicted entities.

4.2 Gesture

Overall, the data contained 185 gestures, of which 67 (36.2 %) were produced by the adult participants, and 118 (63.8%) were produced by the children. Generally, the participants varied with respect to the number of gestures they produced as shown in Figure 23, with five participants producing no gestures (2 adults, 3 children), and three participants (1 adult, 2 children) producing more than two-thirds of the total gestures (69.2%). Therefore, given the small dataset and the variation between participants, the results discussed below should be considered exploratory, as they may be sensitive to individual variation.

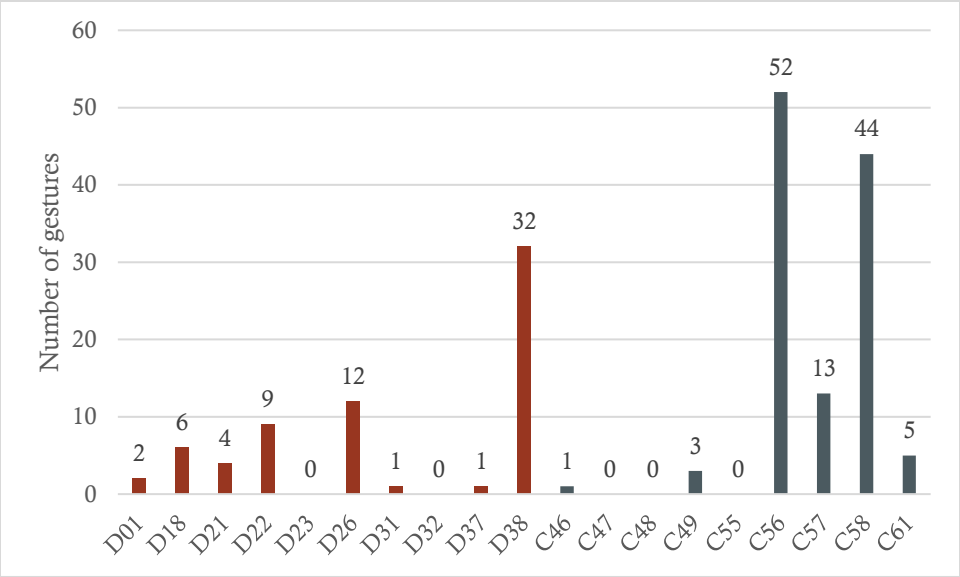


Figure 23 - Gestures produced pr. participant.

Of the total gestures, 115 gestures (62.2%) were coded as iconic. These were used to describe target stimuli more frequently than control stimuli, as shown in Figure 24, while non-iconic gestures were used with the same frequency for control and target stimuli.

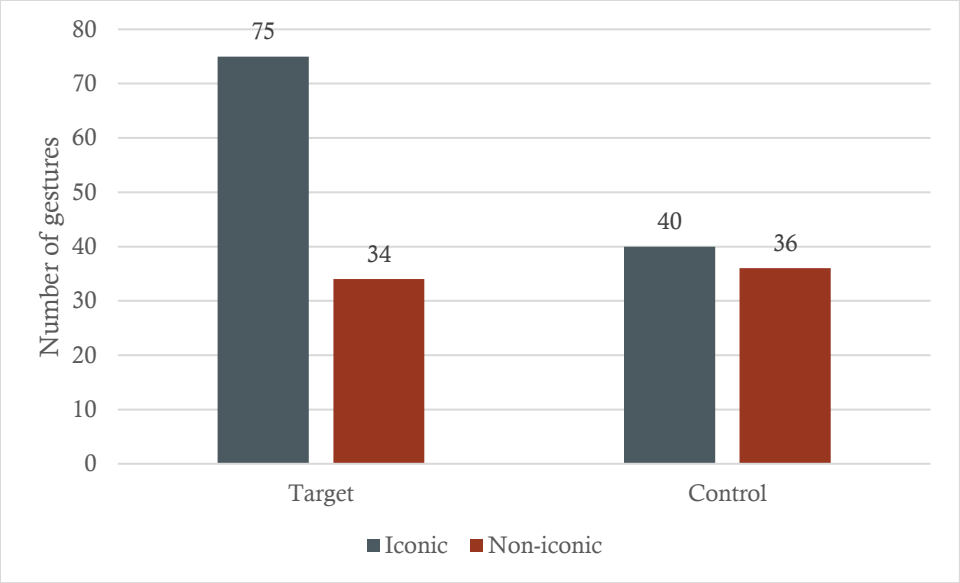


Figure 24 - Distribution of iconic and non-iconic gestures across target and control stimuli.

Looking only at the iconic gestures, Figure 25 shows that NAM-gestures (T, M, MM) were used more frequently for target-stimuli than control-stimuli, whereas the static gestures (SP, E) were used for both with similar frequency, which confirms the conjecture that NAM-gestures were in fact used to express non-actual motion (see Section 3.5 for definitions).

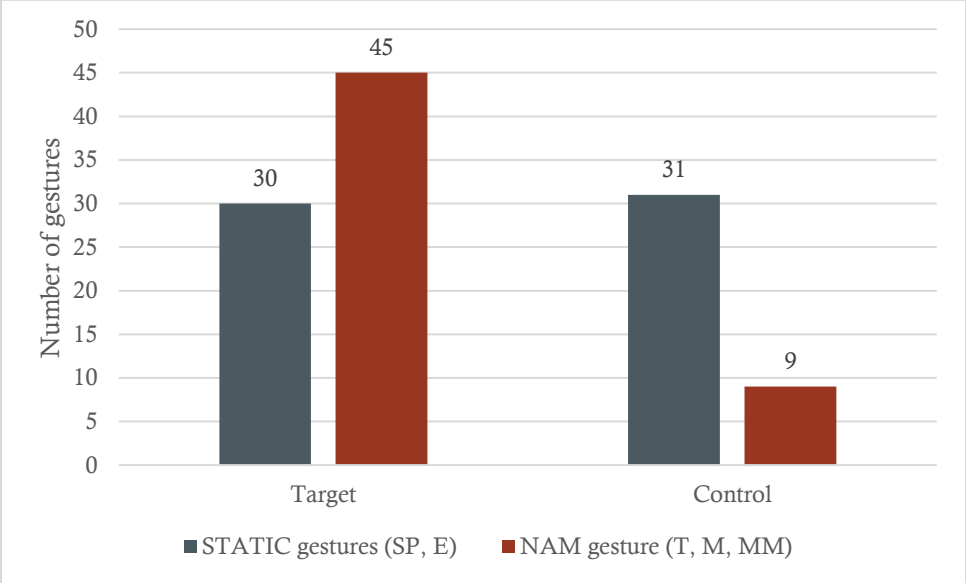


Figure 25 - Distribution of static and NAM-gestures across target and control stimuli.

All four target stimuli types elicited both NAM-gesture and static gestures as shown in Figure 26. While the static gestures were distributed relatively equally across the four target stimulus types, [1pp, +afford] stimuli elicited a slightly higher amount of NAM-gestures than the remaining three kinds of target stimuli, which elicited similar frequencies.

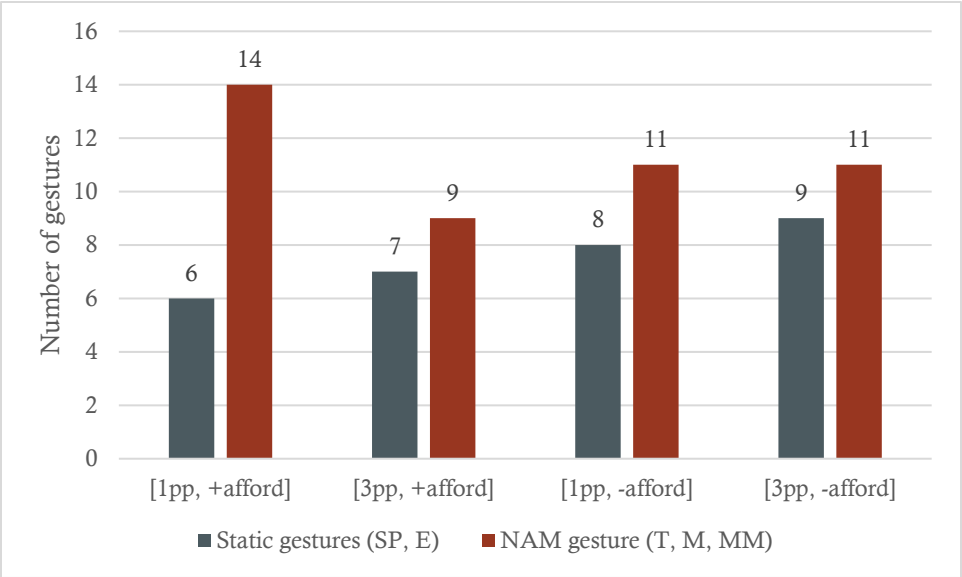


Figure 26 - Distribution of static and NAM-gestures across target stimuli types.

4.2.1 Gesture iconicity-type

Due to their reflections of different experiential motivations, different types of iconic gestures were expected to be produced for different types of stimuli. As seen in Table 7, Motion (M) and Motion-Manner (MM) gestures occurred rarely in the dataset, with five and two occurrences, respectively. This makes it difficult to make generalizations about their usage. The M gestures were elicited by [1pp, -afford] (n=3) and [3pp, -afford] (n=2), and the MM gestures were elicited once each by [3pp, -afford] and [control]. This did not follow the predicted pattern.

Table 7 - Distribution of iconic gesture-types across the stimuli.

	Shape/ placement	**Extension**	**Trace**	**Motion**	**Motion-Manner**
[1pp, + afford]	*3*	*3*	*14*	*0*	*0*
[3pp, + afford]	*0*	*7*	*9*	*0*	*0*
[1pp, -afford]	*4*	*4*	*8*	*3*	*0*
[3pp, -afford]	*2*	*7*	*8*	*2*	*1*
Total Target	9	21	39	5	1
Total Control	11	20	8	0	1
TOTAL	**20**	**41**	**47**	**5**	**2**

Six out of seven M and MM gesture were hand-as-hand gestures, produced by forming a fist around an imaginary object, and running the hand along it, as illustrated in Figure 27. In most cases the difference between M and MM could be taken to be a consequence of the shape of the object, being represented either as straight, or as having more a complex shapes, leading to a non-linear trajectory.

Figure 27 - Example of a hand-as-hand gesture produced in description of chain 3pp.

The majority of NAM-gestures produced were of the Trace kind. These were predicted to be elicited by all four types of target stimuli, which was found to be the case, with the highest frequency for [1pp, + afford], and the remaining target stimulus types eliciting equal amounts, as shown in Figure 28. They were rarely elicited by control stimuli, as shown in Figure 29. The two types of static gestures, SP and E, were both produced with equal frequency for control and target stimuli (see Figure 29). It is interesting to note that extension gestures were produced more frequently for stimuli depicted from the 3pp viewpoint than the 1pp viewpoint (see Figure 28).

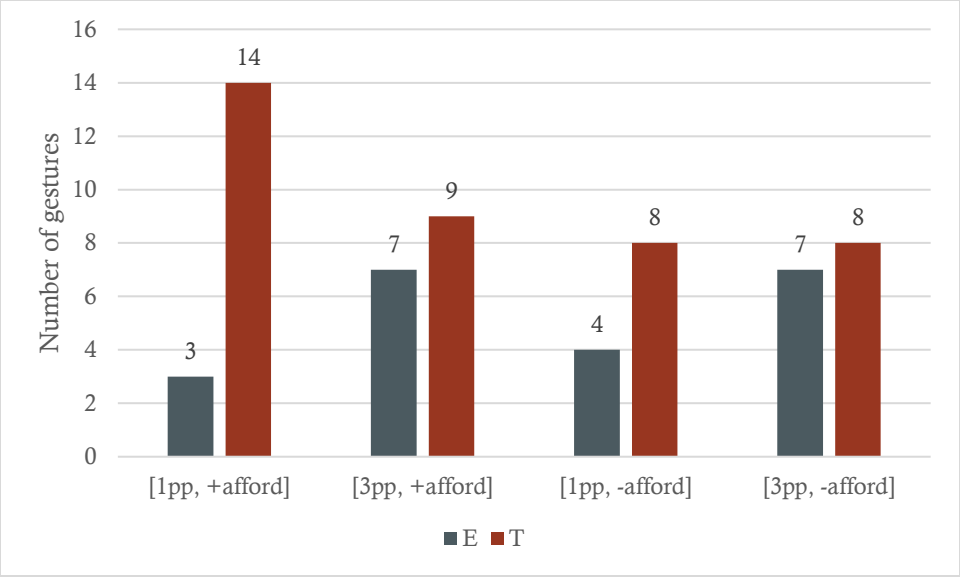


Figure 28 - Distribution of extension and trace gestures across the target stimulus types

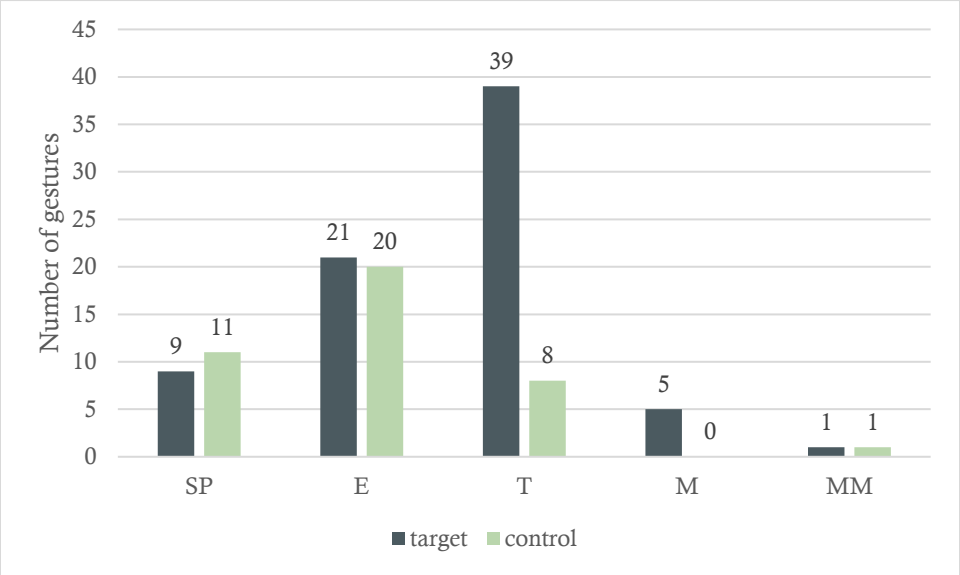


Figure 29 - Distribution of target and control stimuli pr. gesture type.

4.3 Speech-gesture interaction

To explore the interaction between NAM-expressions in speech and gesture, it was tested how many of the NAM-gestures were produced within a picture description containing at least one

linguistic NAM-expression. This was the case for 28 out of 54 NAM-gestures (51.8%). These gestures will henceforth be called *co-occurring NAM-gestures.*[20] The remaining 26 NAM-gestures (48.1%), were produced within descriptions that did not contain a linguistic NAM-expression, henceforth called *freestanding NAM-gestures*. There were considerable differences in the distribution of co-occurring NAM-gestures and freestanding NAM-gestures across the different stimuli types.

As shown in Figure 30, co-occurring NAM-gestures were used almost exclusively to describe target stimuli, while freestanding NAM-gestures were used to describe both target and control stimuli, with approximately one third used to describe the latter.

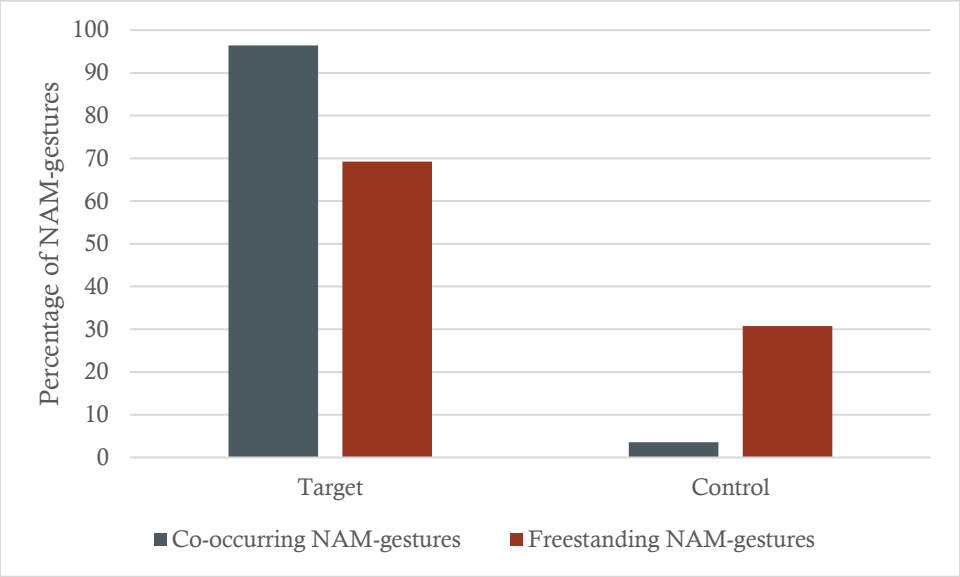


Figure 30 - Proportional distribution of co-occurring and freestanding NAM-gestures across target and control stimuli

Furthermore, there were differences in the distribution of co-occurring NAM-gestures and freestanding NAM-gestures across the four types of target stimuli, as can be seen in Figure 31. Co-occurring NAM-gestures were used more frequently to describe [1pp, +afford] stimuli,

[20] Notice that this measure is not sensitive to whether the NAM-gesture occurs specifically within the NAM-clause.

than for the remaining three types of target stimuli, whereas freestanding NAM-gestures were distributed relatively equally across all four types of target stimuli descriptions.

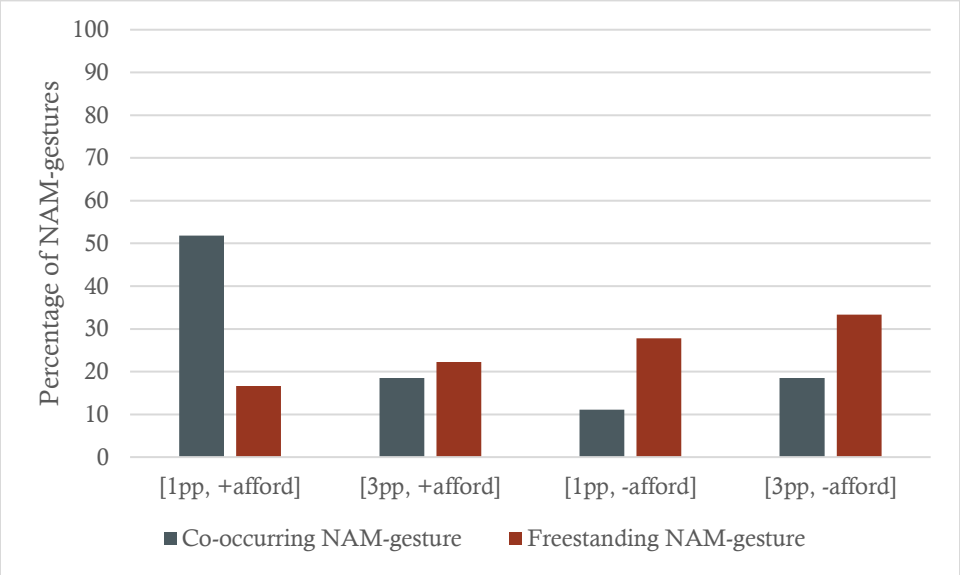


Figure 31 - Proportional distribution of co-occurring and freestanding NAM-gestures across the four types of target stimuli.

NAM-gestures aligned with many different form classes, for both co-occurring and freestanding NAM-gestures. However, for co-occurring NAM-gestures the most frequent alignment was with adverbs or prepositions, as exemplified in (4.8), while for freestanding NAM-gestures it was with nouns, as exemplified in (4.9).

					|--8a--|		|--8b--|	
(4.8)	Trappan	verkar	gå	uppåt	eller	uppför	en backe	
	The stairs	appears	go	upwards	or	up	a hill	
	eller	berg	eller	nåt.				
	or	mountain	or	something				

‘The stairs seems to go upwards, or up, a hill or mountain or something.’

(D38_stairs1pp)

Figure 32 - Trace gesture produced in description of Stairs 1pp (alignment 4.8 8a)

Figure 33 - Trace gesture produced in description of Stairs 1pp (alignment 4.8 8b)

|-9a-|

(4.9) En grotta med en väg in i grottan.
A cave with a road in in the cave

'A cave, with a road into the cave.'

(C56_cave3pp)

Figure 34 - Trace gesture produced in description of Cave 3pp (alignment 4.9 9a)

Another, less frequent, type of speech-gesture alignment was of co-occurring NAM-gestures with motion verbs, possibly together with an adverb/preposition, as in example (4.10). Obviously, freestanding NAM-gestures were not aligned with motion verbs, and only few were aligned with verbs in general, which were either expressing existence or perception, as in example (4.11).

(4.10)	En brygga	som	vi	står	framför	som	att	vi	ska
	A bridge	that	we	stand	in front of	as	that	we	shall
	|----------10a----------|								
	gå	igenom	det.						
	walk	through	it						

'A bridge that we stand in front of, like we are about to walk through it.'

(D38_bridge1pp, repeated from 4.3)

Figure 35 - A trace gesture produced in description of Bridge 1pp (alignment 4.10, 10a)

|----------11a----------|

(4.11)	Man	står	vid	staketet	och	tittar	på	trädet.
	One	stand	by	the fence	and	look	at	the tree

‘One is standing by the fence, looking at the tree.’

(C56_fence1pp)

Figure 36 - A trace gesture produced in description of Fence 1pp (alignment 4.11, 11a)

The analysis of gesture-speech interactions showed that not every instance of NAM-gestures are related to the expression of non-actual motion. Freestanding NAM-gesture as in example (4.12) show that uni-vectorial gestures produced with a neutral handshape may have additional usages, beyond expressing motion.

(4.12)	Det	är	en säng	på	mitten	av	bilden	som
	It	is	a bed	on	the middle	of	the picture	that
			|-12a--|					
	har	blå täcke	eller	överkast				
	has	blue quilt	or	bedspread.				

'There is a bed in the middle of the picture, that has a blue quilt or bedspread.'

(D38_lampBed)

Figure 37 - Trace gesture produced in description of Lamp bed (alignment 4.12, 12a)

4.4 Summary

The results of the linguistic analysis concerning Prediction 1 showed that overall the target stimuli elicited more NAM-Es than control stimuli. The [+afford] stimuli elicited more NAM-Es than the [-afford] stimuli. There was no considerable difference between the amounts of gestures elicited by afford-stimuli depicted from first- and third-person viewpoint, respectively, as had been predicted. This raises some further questions about the role of enactive self-motion and affordance of motion as motivations for expressions; and the connection between the amount of possible motivations for a scene and the tendency to describe that scene using NAM-expressions, which are addressed in Chapter 5. An unpredicted difference was seen in the amounts of NAM-Es elicited by [1pp, -afford] and [3pp, -afford] stimuli, which was suggested to be related to the

description of non-perceivable motion of light-particles. For narrow NAM-expressions, the [+afford] stimuli elicited more N-NAM-Es than [-afford] stimuli, but unlike NAM-expressions broadly, [1pp, +afford] elicited fewer N-NAM-Es than [3pp, +afford].

It was found that extended NAM-expressions were used with equal frequency to describe target and control stimuli. For the target-stimuli, [1pp, +afford] stimuli elicited more X-NAM-Es compared to the remaining three types. Closer analysis showed that X-NAM-Es may be sub-divided in two four types, where two were primarily used to describe target stimuli, expressing either imagined motion of a superimposed subject (I-NAM), or the potential for motion (P-NAM), while two others, expressing results of motions (R-NAM), and (perceived) on-going motion (S-NAM), were primarily used to describe control stimuli.

Taken all together, these results suggest that all three investigated NAM-experiences; visual scanning, affordance of motion, and enactive self-motion, play a role in motivating NAM-expressions to various degrees. The results also support the decision to include not only narrow NAM-expressions, but also extended NAM-expressions, as they appear to play different, but related, roles in the description of static scenes, indicating that both reflect the phenomenon to non-actual motion. The results from the linguistic analysis thus provide partial support for the prediction that [control], [-afford], [3pp, +afford], and [1pp, +afford] would elicit NAM-Es with increasing frequency (P1). With the limitation that no difference was found between stimuli with two- and three possible motivations, the results did show a difference between stimuli with zero, one, and more than one possible NAM-motivation.

As previously stated, the gesture analysis should be considered exploratory, given the low number of gestures in the dataset, and the variation in gesture-frequency between participants. Given these reservations, the gesture data showed that non-iconic and the static types of iconic gestures were used with equal frequency for control and target stimuli, whereas the NAM-gestures were used more frequently for target stimuli, confirming that these gestures are relevant to the phenomenon of non-actual motion. All four types of target stimuli were shown to elicit NAM-gestures, with [1pp, + afford] stimuli eliciting the highest frequency, and the remaining three types eliciting equal amounts. This means that in the case of NAM-gestures, no difference in elicitation was observed between stimuli with one and two possible motivations, However,

stimuli with three possible motivation did elicit more NAM-gestures, and all scenes with possible motivations elicited more NAM-gestures than the control stimuli, with no possible motivations. This may be taken as partial support for Prediction 1 concerning gesture. This difference between speech and gesture, and within speech for N-NAM-expressions and X-NAM-expressions suggest that the relationship between the motivation and expression of non-actual motion varies both across and within semiotic systems, as discussed in the following chapter.

It was predicted that the different types of NAM-gestures would be elicited by different types of stimuli, due to their being linked to different motivations (P2). M and MM gestures occurred rarely in the dataset, and did not follow the predicted pattern, which stated that MM gestures would be elicited by [1pp, +afford] stimuli, and M gestures by [+afford] stimuli. Due to the low number of occurrences, the current data does not allow to evaluate P2a and P2b. As predicted, T gestures were elicited by all four types of stimuli, supporting P2c. However, the fact that [1pp, +afford] stimuli elicited a considerably higher frequency of T gestures, which was hypothesized to be motivated by visual scanning warrants further discussion, which is presented in Chapter 5.

The exploration of speech-gesture interaction showed that NAM-gestures are used both alone and together with linguistic NAM-expressions. This suggests that expressions of non-actual motion is not limited to the semiotic system of language, but may be done either in speech alone, gesture alone, or in gesture and speech together. Co-occurring and freestanding NAM-gestures were equally frequent in the dataset, but there were some interesting differences in their distribution and speech alignment. The co-occurring NAM-gestures were produced almost exclusively for the target stimuli, where it was used to describe [1pp, +afford] stimuli in more than half of the instances, with the remaining occurrences distributing equally across the remaining three types of target stimuli. In comparison, one third of the free-standing NAM-gestures were used to describe control stimuli, and the remaining two-thirds were divided relatively equally across descriptions of all four types of target stimuli. Further it was shown that NAM-gestures occurred in alignment with a range of different parts of speech, but only rarely with verbs, including motion verbs. Instead, co-occurring NAM-gestures aligned primarily with adverbials and prepositions, while free-standing NAM-gestures aligned primarily with nouns.

5 Discussion

In this chapter, I discuss the findings reported in the previous chapter in relation to the general hypotheses, given in Chapter 2. Where it is relevant, I will suggest how these findings may inform future studies. This chapter therefore is a return to the conceptual ("what") side of the conceptual-empirical loop (see Section 2.1.1), and points forward to future explorations on the empirical ("how") side of the loop.

To remind, the hypotheses were:

- **H1** The more NAM-motivations that may apply to a given static scene, the more likely it is to be described with a NAM-expression in speech and gesture.
- **H2** Different NAM-experiences may be reflected by different types of iconic gestures.

The results of the study presented in Chapter 4 indicate that visual scanning, affordance of motion and enactive self-motion are relevant as motivations for NAM-expressions. However, they also indicate that frequency alone is not enough to determine the influence of the different NAM-motivations. Therefore, I turn first to a discussion of how different constructions in language, and different types of gestures may be taken as support for the active role of each of the three investigated NAM-motivations.

5.1 More motivations, more expressions?

5.1.1 Speech

For speech, it was found that scenes associated with two or more possible NAM-motivations led to the production of more expressions of non-actual motion than pictures associated with a single motivation, but there was not, as hypothesized, a difference between pictures with two and three possible NAM-motivations. Given that there was no considerable difference in NAM-E

elicitation by pictures that could invite both visual scanning and affordance of motion and those which could additionally invite enactive self-motion, one might conclude that enactive self-motion does not function as an additional motivation for the linguistic expression of non-actual motion.

However, the analysis of narrow and extended NAM-expressions individually, showed that for extended NAM-expressions, those pictures inviting enactive self-motion elicited the greatest amount of NAM-expressions. This indicates that all three types of NAM-experiences function as motivations for NAM-expressions, but it also shows that there is not, as hypothesized a linear relationship between the number of possible NAM-motivations for a scene, and the frequency with which the scene will be described using a NAM-expression. It will therefore be beneficial to look at the conditions under which each of the three motivations may function, by looking for types of constructions that may be indicative of a specific motivation, and how they may help us understand the relationship between possible motivations and expressions of non-actual motion.

5.1.1.1 Enactive self-motion

It was found that a type of extended NAM-expression (I-NAM, see page 58) was used to describe a scene as if someone was moving through it. In some cases that "someone" was clearly the speaker, who used the personal pronouns *jag* ('I') and *vi* ('we') to express the experience of the stimuli as if they were moving through it, which would be indicative of the enactive self-motion experience as a motivation for the NAM-expression. In some cases, the experience was presented as more generalized, using the impersonal pronoun *man* ('one'). However, it is possible that even these cases are motivated by enactive self-motion of a more generalized kind.

The fact that these expressions were explicitly marked in speech to be "as if" or "like" indicates that the enactive self-motion motivation may be related to imagination, presented in Section 2.3.1.2 as an experience which does not posit existence, but rather constructs it (Blomberg & Zlatev, 2014). In Chapter 2, imagination was introduced as a motivation for methaphorical NAM-expressions, rich in manner-of-motion information, which were not included for systematic elicitation within this study. However, these results indicate that

imagination may play a role in the motivation of non-actual motion expressions beyond metaphoricity.

5.1.1.2 Affordance of motion

Another type of extended NAM-expression (P-NAM, see page 58), was used to describe the typical usage of an object, which may be linked to the experience of affordance of motion. As this type of expression was used in descriptions of control stimuli as well as target stimuli, it may possibly be linked to the experience of affordances more generally, not only affordance of motion.

It was also found that the second most frequent verb in narrow NAM-expressions, *leda* ('lead') was used primarily in descriptions of objects that afford motion, which suggests that the experience of affordance of motion may be expressed both in certain narrow and certain extended NAM-expressions.

5.1.1.3 Visual scanning

There were no examples in the data of specific linguistic constructions which could clearly be linked to the experience of visual scanning. However, the finding that pictures of elongated, static objects that do not afford motion consistently elicited N-NAM-Es indicate the role of visual scanning as a relevant motivation for expressions of non-actual motion, as it is hard to argue that these could be experientially motivated by any of the other NAM-experiences.

We could of course consider that the description of static situations through the use of motion verbs may be motivated by established convention, especially given the "lack" of alternative available constructions in Swedish to describe this kind of stimuli (Blomberg, 2014, p. 172), which would either require extensive explications, as in (5.1), or resorting to simple statements of existence. However, as is the case for semiotic grounds (see Section 2.1.3), conventionality does not entail the absence of other motivations.

> (5.1) The road has a certain configuration with respect to the forest: the initial part (closest to us) is outside, the further part (away from us) is inside…
>
> (Blomberg & Zlatev, 2014, p. 412)

5.1.2 Gesture

It was found that iconic gestures were used more frequently to describe target stimuli than control stimuli, whereas non-iconic gestures were used with similar frequency for both. Furthermore, the proportion of NAM-gestures used to describe target stimuli was considerably greater than for control stimuli, whereas static gestures were used with the same frequency for both. This indicates that NAM-gestures have an affinity with the types of stimuli designed specifically to elicit NAM-expressions. Thus, we may indeed consider the usage of these gestures as gestural expressions of non-actual motion.

As for speech, it was found that there was not, as hypothesized, a linear relationship between the number of possible NAM-motivations for a scene, and the number of NAM-gestures elicited in the descriptions of the scene. Rather, it was found that all target picture types elicited NAM-gestures, but that pictures which may invite visual scanning, affordance of motion and enactive self-motion elicited the greatest amount. The finding that there was no considerable difference in the amount of NAM-gestures elicited by pictures which may only invite visual motion, and pictures which may invite both visual scanning and affordance of motion may suggest that the experience of affordance of motion does not function as a motivation for gestural expression of non-actual motion. This might be a reasonable interpretation, given that affordance of motion is a rather "abstract" NAM-motivation, in the sense that it is not the experience of motion, but rather the experience of potential for motion. While language has the option to mark such potentials, as was done for example in X-NAM-Es with the use of the phrase *som man använder till* ('that one use for'), the semiotic system of gesture may not have any immediate resources to mark the difference between potential and concrete actions, suggesting that it could be less apt for expressing experiences of the affordance of motion.

5.2 Different experiences, different gestures?

It was hypothesized that different NAM-experiences would be reflected in different types of NAM-gestures, based on an iconic relationship between the NAM-experience and the features of the gesture.

The types of stimuli included in this study did not elicit many Motion and Motion-Manner gestures, meaning that there is not sufficient data to establish any clear patterns of usage for these gestures. It may be the case that these are not generally enlisted for the expression of non-actual motion, or it may be the case that something about the communicative situation was not conductive to their production. We may gain some insight from the fact that most Motion and Motion-Manner gestures were produced with the hand-as-hand configuration, which suggests that in these limited cases, these gestures were produced to represent movements of some human entity, possibly the speakers themselves, while only one hand-as-object gesture was produced, indicating that the speakers generally did not use gestures to represent the movement of entities depicted in the picture.

With the expression of self-motion seen in linguistic X-NAM-Es, one might have expected to find more enactive self-motion gestures, expressing the movement of one's own body through the scene. However, the fact that participants were seated, and most relevant movements like walking on a road, climbing on a ladder and so on are whole-body movements, may account for the absence of such gestures in the data. To this end, it may be mentioned that the Motion-Manner gestures that did occur, were of actions performed with the hands alone, for example spilling or pouring content out of a vase.

Trace gestures were considerably more frequent in the data. These gestures were used in description of all four kinds of target stimuli, and as such, may indeed be motivated by visual scanning. However, the finding that they occurred more frequently for the type of stimuli that invites enactive self-motion may indicate that this type of gesture may have a more complex rooting in the NAM-motivations than hypothesized. It may be that the increased production of Trace gestures for stimuli that invite enactive self-motion is due to an inherent connection between self-motion and perception (as discussed in Section 2.3.1.3). To walk down a road we

must not only move our body, but also continuously assess the path ahead, which suggests that enactive self-motion and visual scanning may be strongly linked. It might then be the case that Trace gestures can be used either to represent the dynamic process of gaze alone, or of gaze and the individual simultaneously. Alternatively, it may be the case that Trace gestures are linked not to visual scanning itself, but to the expression of Path information, in which case it could either be used to represent the (abstract) path of the gaze (a case of *noesis*), or the concrete path of a real or imagined entity (a case of *noema*).

5.3 Summary

Based on the current discussion, it has become evident that the relationship between NAM-experiences and NAM-expressions is not a simple one. However, the findings of the thesis may hopefully serve as a step towards a richer understanding of the experiential motivation of NAM-expressions, as it was observed that not all NAM-motivations may be equally potent within and across the semiotic systems of language and gesture. Thus, the findings of this thesis suggest that visual scanning may be reflected in narrow NAM-expressions and Trace-gestures, affordance of motion: in narrow and extended NAM-expressions, and enactive self-motion: in extended NAM-expressions, and possibly Trace-gestures.

6 Summary and conclusions

This thesis has explored the phenomenon of non-actual motion (NAM) and its expression in language and gesture, with the aim of answering three research questions, here repeated from Chapter 1:

- **RQ1** What are the possible experiential motivations for NAM-expressions in language and gesture?
- **RQ2** How do these motivations influence the realizations of linguistic and gestural expression of NAM (in Swedish)?
- **RQ3** How do gesture and speech interact in the expression of NAM?

With a basis in NAM theory (Blomberg, 2014; Blomberg & Zlatev 2014), this thesis distinguished between the *experience* of non-actual motion and the *expression* of non-actual motion, where the first is proposed to motivate the latter. Based on phenomenological analysis, three distinct NAM-experiences have previously been proposed as possible motivations for non-actual motion expressions; *visual scanning*, as a dynamic process of perception, *imagination*, as an active process of constructing experience and *enactive perception*, as the directed process of attending to objects as potentials for action (Blomberg, 2014). In this thesis, I proposed that enactive perception be further distinguished into *affordance of motion* and *enactive self-motion*, where the first may be understood as a generalized perception of affordances, while the latter is strongly connected to a subjective, enactive experience of agency.

A review of the existing literature showed that a wide range of typologically varied languages employ words with motion-semantic meanings in description of static configurations, which is the general definition of NAM-expressions (Blomberg, 2014). It was seen that languages differ in the specific conventions governing the use of such expressions (Blomberg, 2014; Matsumoto, 1996), but translation studies also showed that the motion-semantic component of NAM-expressions were retained in translations (Rojo & Valenzuela, 2003; Stosic

& Sarda, 2009) suggesting that their experiential motivation is not completely conventionalized (sedimented).

In the present analysis of NAM-expression in language (realized here as speech), NAM-expressions were broadly defined as *an expression using a motion verb to describe a static scenario*, and a distinction was made between narrow NAM-expressions, where the grammatical subject of the motion verb is inanimate and the motion-semantic meaning cannot be interpreted to describe actual motion, a definition that has been employed in previous empirical research (Blomberg, 2014). The present thesis included additional expressions containing a motion verb used to describe static scenarios under the collective term *extended NAM-expressions*. Further, extended NAM-expressions were grouped into four categories: I-NAM expressing imagined motion of an imposed subject, P-NAM expressing typical usage of depicted objects, R-NAM, describing static situations as the result of motion, and S-NAM describing static scenarios as a "frozen" moment of an ongoing motion.

As a further novel aspect, this thesis also studied the gestures produced in descriptions of static scenarios. Based on formal and semantic features of the gesture, a distinction was made between static gestures and NAM-gestures, with the latter including three gesture types: Trace, Motion and Motion-Manner gestures.

An empirical study was conducted to explore the relationship between the possible NAM-motivations and the expression of non-actual motion in speech and gesture. Participants were presented with drawings depicting elongated, static objects, which either afforded human motion, such as roads, paths and ladders, or did not afford human motion, such as fences, chains and cracks, presented from both a first- and third-person viewpoint, together with control drawings depicting static objects with little or no elongation, such as apples, benches, and vases. The design of the stimuli was systematically varied to favor different NAM-motivations, with depictions of traversable objects from a first-person perspective inviting the experience of visual scanning, affordance of motion and enactive perception, traversable objects from a third-person perspective inviting visual scanning and affordance of motion, and depictions of non-traversable objects, regardless of perspective inviting visual scanning. The control stimuli were designed to

not invite any NAM-experiences. It was hypothesized that the more possible NAM-motivations a picture invited, the more NAM-expressions it would elicit.

The findings of the study suggest that there is not a linear relationship between the number of NAM-motivations and the frequency of NAM-expressions. Rather, the relationship between NAM-motivations and NAM-expressions appear to governed in part by the number of possible NAM-motivations, in part by the extent of resources available to express the different NAM-experiences in the semiotic systems of language and gesture.

In speech it was found that visual scanning was reflected in narrow NAM-expressions, enactive self-motion in extended NAM-expressions and affordance of motion in both narrow and extended NAM-expressions. This may help explain why there was no difference in the amount of NAM-expressions produced in descriptions of traversable objects depicted from a first- and third-person perspective. This finding suggests that while the different NAM-motivations may be based on general cognition, the potential for their expression may be language dependent, suggesting that there could be differences between languages in the amount of NAM-expressions that different scenes may elicit. Future research should therefore compare the production of NAM-expressions in different languages, beyond previous work (Blomberg, 2014).

For gestural expression of NAM, it was found that there was no difference in the frequency of NAM-gestures elicited by pictures which invite only visual scanning, and pictures which invite visual scanning and affordance of motion, while there was a considerable difference between these, and pictures which additionally invite enactive self-motion. This may be taken to suggest that affordance of motion does not motivate gestural expressions of non-actual, and it was proposed that this is due to a lack of possibility to mark the difference between specific instances of motion and general potentials for motion in gestures.

Three different types of iconic gestures were proposed as candidates for gestural expression of NAM. Only one of these, Trace gestures, was frequently occurring in the dataset. Trace gestures were used to describe both pictures that invite visual scanning alone, and those that additionally invited affordance of motion and enactive self-motion. It was therefore suggested that Trace gesture may either co-express visual scanning and self-motion, given that

these are inherently linked. Alternatively, it may be taken to express path of movement more generally, be it of the gaze or of a moving entity.

The remaining two gesture types proposed, Motion and Motion-Manner gestures, occurred rarely in the dataset. This may suggest that either these gestures are not used in the expression of non-actual motion, or the communicative situation revolving the experiment was not accommodating to the expression of these gestures. The latter might be especially relevant for Motion-Manner gestures, which are likely to be produced with the whole body. Considering the small dataset used in this study, more research is needed to determine the role of Motion gestures and Motion-Manner gestures in the gestural expression of non-actual motion.

In sum, this thesis has established that non-actual motion expressions are not a homogeneous linguistic phenomenon, nor is its expression limited to the semiotic system of language. It was found that there is a relationship between NAM-motivations and NAM-expressions, but that this is complex, and influenced not only by the range of NAM-experiences applicable to a certain scene, but also by the linguistic and gestural resources available to express this experience.

These findings underline the importance of distinguishing non-actual motion as an experiential phenomenon from the way it is expressed in language and gesture, in their respective NAM-expressions. A clear separation between experience and expression enables a systematic exploration of the role of NAM-experiences as motivations for NAM-expressions, which underscores the importance of phenomenological analysis in cognitive semiotics. The present thesis has contributed to NAM-theory, and to cognitive semiotics in general, with the proposal of a new, distinct NAM-motivation: enactive self-motion. On the expression side, it has contributed by showing that the semiotic system of gesture, as well as that of language, is involved in the expression non-actual motion. Lastly, the distinction between narrow and broad NAM-expressions proposed has led to new insights into the types of resources available for expression of non-actual motion in language, which will be especially interesting for future cross-linguistic research on the topic.

References

Amagawa, T. (1997). Subjective motion in English and Japanese: A case study of run and hashiru. *Tsukuba English studies*, *16*, 33-50.

Andersen, L. (2018). *Good hand, bad hand: correlations between gesturing hand and speech valence*. [B.A. thesis]. Retrieved from: http://lup.lub.lu.se/student-papers/record/8959153

Andrén, M. (2010). *Children's Gestures from 18 to 30 Months.* Lund: Centre for Languages and Literature, Lund University.

Blomberg, J. (2014). *Motion in language and experience: actual and non-actual motion in Swedish, French and Thai*. Lund University.

Blomberg, J. (2015). The expression of non-actual motion in Swedish, French and Thai. *Cognitive Linguistics*, *26*(4), 657-696.

Blomberg, J. (2017). Non-actual motion in language and experience. *Motion and Space across Languages: Theory and applications*, *59*, 205-228.

Blomberg, J., & Zlatev, J. (2014). Actual and non-actual motion: Why experientialist semantics needs phenomenology (and vice versa). *Phenomenology and the cognitive sciences*, *13*(3), 395-418.

Blomberg, J., & Zlatev, J. (2015). Non-actual motion: phenomenological analysis and linguistic evidence. *Cognitive processing*, *16*(1), 153-157.

Brandt, L. (2009). Subjectivity in the act of representing: The case for subjective motion and change. *Phenomenology and the cognitive sciences 8*(4), 573–601.

Brown, S., Mittermaier, E., Kher, T., & Arnold, P. (2019). How pantomime works: implications for theories of language origin. *Front. Commun*, *4*(9), 1-17.

Bundgaard, P. F. (2010). Husserl and language. In *Handbook of phenomenology and cognitive science* (pp. 368-399). Springer.

Capirci, O., Cristilli, C., De Angelis, V., Graziano, M., Stam, G., & Ishino, M. (2011). Learning to use gesture in narratives: developmental trends in formal and semantic gesture competence. *Integrating gestures: the interdisciplinary nature of gesture*, 189-200.

Cartmill, E. A., Rissman, L., Novack, M. A., & Goldin-Meadow, S. (2017). The development of iconicity in children's co-speech gesture and homesign. *Language, Interaction and Acquisition*, *8*(1), 42-68.

Chandler, D. (2007). *Semiotics: the basics*. Routledge.

Coseriu, E. (1977). Linguistic (and other) universals (Valerie Becker Makkai, Trans.). In A. Makkai, V.B. Makkai, & L. Heilmann (Eds.), *Linguistics at the Crossroads* (pp. 317-346). Padova: Jupiter press.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston, MA.

Green, J. (2014). *Drawn from the ground: Sound, sign and inscription in Central Australian sand stories (No. 13)*. Cambridge University Press.

Gullberg, M. (2011). Thinking, Speaking and Gesturing about Motion in more than One. *Thinking and speaking in two languages*, *77*, 143.

Gullberg, M., Hendriks, H., & Hickmann, M. (2008). Learning to talk and gesture about motion in French. *First Language*, *28*(2), 200-236.

Heine, B., & Kuteva, T. (2002). On the Evolution of Grammatical Forms. In A. Wray (Ed.), *The transition to language*, *2* (pp. 376-397) Oxford: Clarendon.

Hickmann, M., Hendriks, H., & Gullberg, M. (2011). Developmental perspectives on the expression of motion in speech and gesture: A comparison of French and English. *Language, interaction and acquisition*, *2*(1), 129-156.

Husserl, E. (1900-1901). *Logische Untersuchungen, 2nd edition*. Tübingen: Max Neimauyer Verlag. (1913/1980). English translation: Logical investigations. London/New York: Routhledge, 2001

Husserl, E. (1973). *Experience and Judgement*. Trans. J. S. Churchill, & K. Ameriks. Evanston: Routledge.

Itkonen, I. (2008). Concerning the role of consciousness in linguistics. *Journal of Consciousness Studies*, *15*(6), 15-33.

Iverson, J. M., & Goldin-Meadow, S. (2005). Gesture paves the way for language development. *Psychological science*, *16*(5), 367-371.

Kendon, A. (1980). Gesticulation and speech: Two aspects of the Process of Utterance. In M. R. Key (Ed.), *The relationship of verbal and nonverbal communication* (pp. 207-228). The Hague: Mouton Publishers.

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

Kita S., van Gijn I., van der Hulst H. (1998) Movement phases in signs and co-speech gestures, and their transcription by human coders. In: Wachsmuth I., Fröhlich M. (Eds.), *Gesture and Sign Language in Human-Computer Interaction. GW 1997. Lecture Notes in Computer Science, vol 1371* (pp. 23-35). Springer, Berlin, Heidelberg

Langacker, R. W. (1987). Nouns and verbs. *Language*, 53-94.

Langacker, R. W. (2006). Subjectification, grammaticization, and conceptual archetypes. In A. Anthanasiadou, C. Canakis, & B. Cornillie (Eds.), *Subjectification: Various paths to subjectivity* (pp. 17-40). Berlin, New York: Mouton de Gruyter.

Matlock, T. (2004). The conceptual motivation of fictive motion. In G. Radden and K-U. Panther (Eds.), *Studies in Linguistic Motivation* (pp. 221-248). Berlin, New York: Mouton de Gruyter.

Matlock, T. (2010). Abstract motion is no longer abstract. *Language and Cognition*, *2*(2), 243-260.

Matsumoto, Y. (1996). Subjective motion and English and Japanese verbs. *Cognitive Linguistics (includes Cognitive Linguistic Bibliography)*, *7*(2), 183-226.

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. University of Chicago press.

McNeill, D. (2005). *Gesture and thought*. University of Chicago Press

Moran, D. (2005). *Husserl: Founder of Phenomenology*. Cambridge: Polity Press.

Müller, C. (2014). Gestural modes of representation as techniques of depiction. In C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, S. Tessendorf (Eds.), *Body–language–communication: An international handbook on multimodality in human interaction*, *vol 2*, (pp. 1687-1702). Berlin, Boston: De Gruyter Mouton.

Naidu, V., Zlatev, J., Duggirala, V., Van De Weijer, J., Devylder, S., & Blomberg, J. (2018). Holistic spatial semantics and post-Talmian motion event typology: A case study of Thai and Telugu. *Cognitive Semiotics*, *11*(2), 1-27.

Özyürek, A., & Kita, S. (1999). Expressing manner and path in English and Turkish: Differences in speech, gesture, and conceptualization. In *Twenty-first Annual Conference of the Cognitive Science Society* (pp. 507-512). Erlbaum.

Özyürek, A., Kita, S., Allen, S., Brown, A., Furman, R., & Ishizuka, T. (2008). Development of cross-linguistic variation in speech and gesture: Motion events in English and Turkish. *Developmental psychology*, *44*(4), 1040.

Peirce, C. S. (1894). What is a sign. Retrieved from: https://www.marxists.org/reference/subject/philosophy/works/us/peirce1.htm

Psychology Software Tools, Inc. [E-Prime 3.0]. (2016). Retrieved from https://www.pstnet.com.

Richardson, D., & Matlock, T. (2007). The integration of figurative language and static depictions: An eye movement study of fictive motion. *Cognition*, *102*(1), 129-138.

Rojo, A., & Valenzuela, J. (2003). Fictive motion in English and Spanish. *International Journal of English Studies*, *3*(2), 123-150.

Rojo, A., & Valenzuela, J. (2009). Fictive motion in Spanish: Travellable, non-travellable and path-related manner information. *Trends in cognitive linguistics: Theoretical and applied models*, 221-239.

Slobin, D. I. (1996). From “thought and language” to “thinking for speaking”. In J. J. Gumperz and S. C. Levinson (Eds.), *Rethinking linguistic relativity* (pp. 70-96). Cambridge University Press.

Slobin, D. I. (2004). The many ways to search for a frog: linguistic typology and the expression of motion events. In S. Strömqvist and L. Verhoeven (Eds.), *Relating Events in Narrative: Typological and Contextual Perspectives* (pp. 219-257). Mahway: Lawrence Erlbaum Associates.

Sokolowski, R. (2000). *Introduction to phenomenology*. Cambridge university press.

Sonesson, G. (1992). *Bildbetydelser: inledning till bildsemiotiken som vetenskap*. Lund: Studentlitteratur.

Sonesson, G. (1997). The ecological foundations of iconicity. In: Rauch, Irmengard; Carr, Gerald F. (eds.), *Semiotics around the world: Synthesis in diversity. Proceedings of the Fifth International Congress of the IASS, Berkeley, June 12– 18, 1994* (pp. 739–742). Berlin and New York: Mouton de Gruyter.

Sonesson, G. (2010). From mimicry to mime by way of mimesis: Reflections on a general theory of iconicity. *Σημειωτκή-Sign Systems Studies, 38*(1-4), 18-66.

Sonesson, G. (2012). The phenomenological road to cognitive semiotics. *Proceedings of the 10th world congress of the International Association for Semiotic Studies (IASS/AIS)*, 855-866.

Stampoulidis, G., Bolognesi, M., & Zlatev, J. (2019). A cognitive semiotic exploration of metaphors in Greek street art. *Cognitive Semiotics*, *12*(1) 1-20.

Stosic, D., & Sarda, L. (2009). The many ways to be located: The role of fictive motion in the expression of static location. In M. B. Vukovic, & L. G. Grmusa (Eds.), *Space and time in language and literature*, (pp. 39-60). Cambridge Scholars Publishing.

Talmy, L. (2000a). *Toward a Cognitive Semantics – Volume 1: Concept structuring systems.* Cambridge, MA: MIT Press.

Talmy, L. (2000b). *Toward a Cognitive Semantics – Volume 2: Typology and process.* Cambridge, MA: MIT Press.

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. (2006). ELAN: A professional framework for multimodality research. In N. Calzolari, K. Choukri, A. Gagemi, B. Maegaard, J. Mariani, J. Odijk, & D. Tapias (Eds.), *Proceedings of the Fifth International Conference on Language Resources and Evaluation* (pp. 1556–1559). Genoa, Italy: European Language Resources Association.

Zahavi, D. (2003). *Husserl's phenomenology*. Stanford University Press.

Zlatev, J. (1997). *Situated embodiment: Studies in the emergence of spatial meaning*. Stockholm: Stockholm University

Zlatev, J. (2009). The semiotic hierarchy: Life, consciousness, signs and language. *Cognitive Semiotics*, *4*(Supplement), 169-200.

Zlatev, J. (2015a). Cognitive semiotics. In P. P. Trifonas (Ed.), *International handbook of semiotics* (pp. 1043-1067). Springer, Dordrecht.

Zlatev, J. (2015b). The emergence of gestures. In B. MacWhinney, & W. O'Grady (Eds.), *The handbook of language emergence* (pp. 458-477). New York: Wiley.

Zlatev, J. (2018). Meaning making from life to language: The semiotic hierarchy and phenomenology. *Cognitive Semiotics*, *11*(1), 1-18.

Zlatev, J. (2019). Mimesis theory, learning, and polysemiotic communication. In M. A. Peters (eds.), *Encyclopedia of educational philosophy and theory*. Singapore: Springer Science and Business Media

Zlatev, J., & Blomberg, J. (2019). Norms of language: What kinds and where from? Insights from phenomenology. In A. Mäkilähde, V. Leppänen, & E. Itkonen (Eds.), *Normativity in Language and in Linguistics* (pp. 69-101). Amsterdam: John Benjamins publishing company.

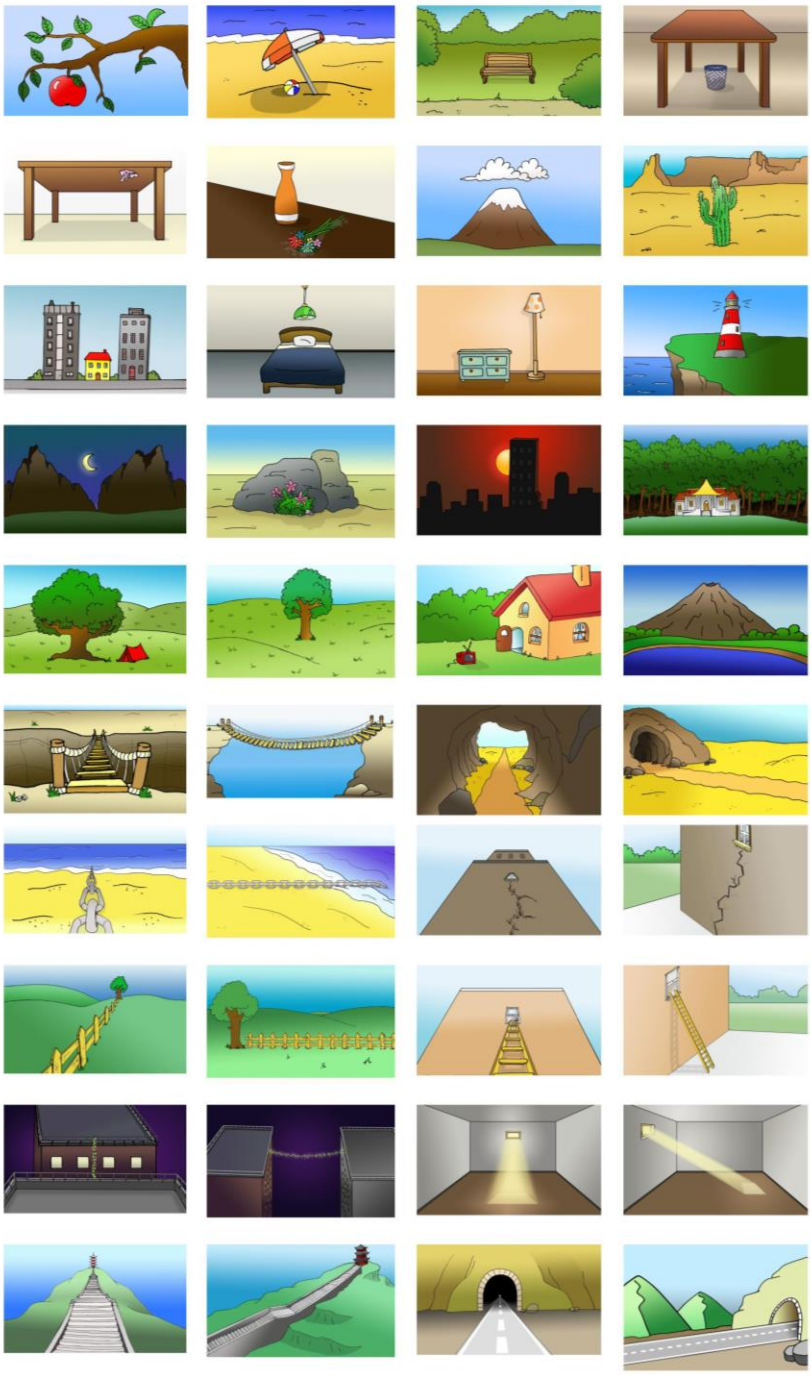
Zlatev, J., Blomberg, J., & David, C. (2010). Translocation, language and the categorization of experience. In V. Evans, & P. Chilton (Eds.), *Language, Cognition and Space: The state of the art and new directions* (pp. 389-418). London, Oakville: Equinox Publishing

Zlatev J., Blomberg, J., Devylder, S., Naidu, V., & van der Weijer, J. (submitted). Motion event description in Swedish, French, Thai and Telugu: For distinct typological patterns.

Zlatev, J., & Yangklang, P. (2004). A third way of travel: the place of Thai in motion-event typology. In S. Strömqvist and L. Verhoeven (eds.), *Relating Events in Narrative: Typological & Contextual perspectives* (pp. 159-190). Mahway: Lawrence Ernbaum Associates.

Zlatev, J., Żywiczyński, P., & Wacewicz, S. (submitted) Pantomime as the original human-specific semiotic system.

Appendix A



Control stimuli

Target stimuli

Appendix B

Figure B1

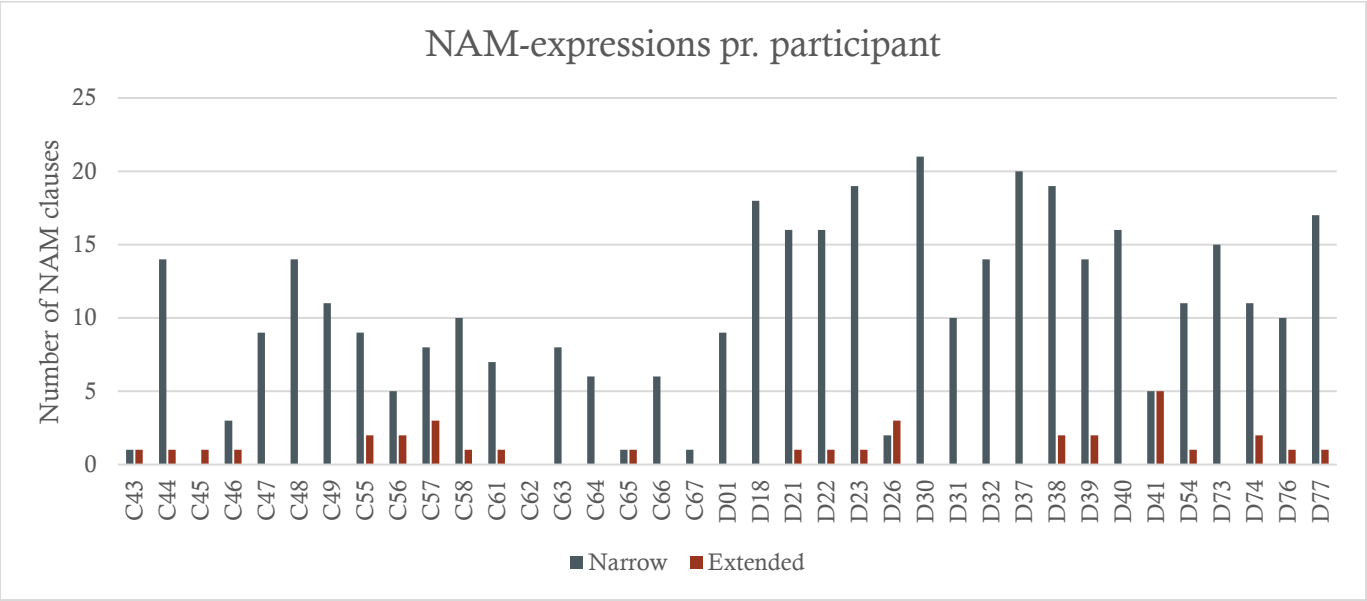


Figure B2