Alevtina Roshchina

# Speech-entangled gesture as a linguistic phenomenon in its own right: towards a multimodal approach to language and evolution

Master’s thesis in M.Phil. English Linguistics and Language Acquisition
Supervisor: Giosuè Baggio
November 2023

Master’s thesis

NTNU
Norwegian University of Science and Technology
Faculty of Humanities
Department of Language and Literature

NTNU
Norwegian University of
Science and Technology

Alevtina Roshchina

# Speech-entangled gesture as a linguistic phenomenon in its own right: towards a multimodal approach to language and evolution

Master’s thesis in M.Phil. English Linguistics and Language Acquisition
Supervisor: Giosuè Baggio
November 2023

Norwegian University of Science and Technology
Faculty of Humanities
Department of Language and Literature

NTNU
Norwegian University of
Science and Technology

# Abstract

This thesis investigates the role of manual(-visual) modality, specifically co-speech gestures, in the context of human communication. While spoken language has traditionally been the primary focus of linguistic research, there is a growing interest in understanding the significance of non-linguistic behaviors and how different modes of communication interact. The study examines how spoken language and co-speech gestures relate to one another, exploring their potential to convey meaning. The work discusses the place of manual(-visual) modality in language theory. It introduces the Multimodal Parallel Architecture (PA) as a theoretical model that can adapt to the dynamic interaction between perception and spoken and gestural modes of expression. Special attention is paid to the discussion about Spatial Structure because our ability to describe and communicate our experiences relies on the integration of mental representation of language and perception. The thesis emphasizes that structural similarities between verbal language and gestures may have arisen from shared cognitive principles or the historical coexistence and mutual reinforcement of these modes. The strength of the framework of Multimodal PA is claimed to be its flexibility and interpretability. In one interpretation, the multimodal approach supports the idea that gestures have their own generative principles for creating semantic and syntactic structures. In this view, gestures are seen as having their own system for conveying meaning, separate from spoken language. This ability also allows co-speech gestures to bypass spoken language and even function independently in some instances. On the other hand, the framework does not fail if gestures are not elaborate enough to generate autonomous structures but can only be integrated into the system established by spoken language. In sum, the thesis shows that adopting a multimodal approach to language and developing both theoretical and experimental basis for future research is a way to answer the questions regarding language evolution and human language faculty.

# Sammendrag

Denne masteroppgaven undersøker rollen til manuell(-visuell) modalitet, spesielt gjester (co-speech gestures), i sammenheng med menneskelig kommunikasjon. Mens muntlig språk tradisjonelt har vært hovedfokuset i lingvistisk forskning, er det økende interesse for å forstå betydningen av ikke-lingvistisk atferd og hvordan ulike kommunikasjonsmodaliteter samhandler. Studien undersøker hvordan muntlig språk og medtankegestikulering forholder seg til hverandre og utforsker deres potensiale for å formidle mening. Oppgaven diskuterer posisjonen til manuell(-visuell) modalitet i språkteorien. Det introduserer Multimodal Parallel Architecture (PA) som en teoretisk modell som kan tilpasse seg det dynamiske samspillet mellom persepsjon og muntlige og gestiske uttrykksformer. Det legges spesiell vekt på diskusjonen om romlig struktur, fordi vår evne til å beskrive og kommunisere våre opplevelser avhenger av integrasjonen av mentale representasjoner av språk og persepsjon. Oppgaven legger vekt på at strukturelle likheter mellom muntlig språk og gjester kan ha oppstått som et resultat av delte kognitive prinsipper eller historisk sameksistens og gjensidig forsterkning av disse modalitetene. Styrken ved rammeverket til Multimodal PA hevdes å være fleksibilitet og tolkningsdyktighet. I en tolkning støtter den multimodale tilnærmingen ideen om at gjester har egne genererende prinsipper for å skape semantiske og syntaktiske strukturer. I dette tilfellet har gjester sitt eget system for å formidle mening, separat fra muntlig språk. Denne evnen tillater også gjester å omgå muntlig språk og til og med fungere uavhengig i noen tilfeller. På den annen side mislykkes ikke rammeverket hvis gjester ikke er tilstrekkelig utviklet til å generere autonome strukturer, men bare kan integreres i systemet etablert av muntlig språk. Samlet sett viser avhandlingen at å ta i bruk en multimodal tilnærming til språk og å utvikle både teoretisk og eksperimentelt grunnlag for fremtidig forskning, er en måte for å besvare spørsmål om språkevolusjon og den menneskelige språkevne.

# Acknowledgements

I would like to express my sincere gratitude to my supervisor, Professor Giosuè Baggio, for his continuous support and guidance throughout my thesis. I am thankful for his interest, encouragement, and efforts in allowing me to explore my ideas while keeping me on track.

In addition to my supervisor, I would like to extend my appreciation to all the members of the Department of Language and Literature (ISL) I had the opportunity to learn from and work with over the last 2.5 years.

On a personal note, I want to express my heartfelt thanks to my mother, Gaurav Chaudhary, Viktoria Afoian, Julia Senderska, and Jon Berg Jørgensen for being my support team throughout this journey and offering helpful advice. Although they likely didn't require detailed, frequent reports of both my excitements and struggles, they patiently listened, and I am beyond grateful for that.

# Table of Contents

# List of Figures

# 1 Introduction

In recent years, researchers within the field of linguistics started recognizing the importance of revisiting the concept of language to account for the variety of semiotic resources employed by speakers when engaged in the process of communication. Although spoken[1] language often remains the primary source of meaning, which helps recognize and correctly interpret the speaker's communicative intent, most of the utterances are multimodal by nature. According to Kendon (2004, p. 127), "the speaker creates an ensemble in which gesture and speech are employed together as partners in a single rhetorical enterprise". Thus, the utterance is no longer seen as a mere sequence of linguistic signs but as "a complete unit of social action which always has multiple components (…) and whose interpretation draws on both conventional and non-conventional signs, joined indexically as wholes" (Enfield, 2013, p. 703). This claim is derived from the general heuristic of semiotic theory (pragmatic unity heuristic), which states that when several signs are produced together, they should be seen as a whole and interpreted as a single composite unit. Moreover, following the contextual association heuristic, if the signs are contextually related, they should be treated as a part of a single signifying action. As a result, the notion of "languaging" or "doing language" comes to the foreground, according to which language should be treated as a complex process where "meaning making" happens through the interplay of various semiotic expressions.

Most of the sign systems simultaneously employ different representational formats. This variability of resources is reflected in Clark's (1996) theory of language use, according to which the meaning of the utterance results in an interplay of three methods of signaling: describing, demonstrating, and indicating. The theory draws on Peirce's triadic theory of the sign (Peirce, 1955) and the notion of "mixed signs", which can display iconic, indexical, and symbolic properties at the same time. In Clark's theory, these notions are extended to the level of the whole linguistic utterance. In other words, language can be characterized by its ability to combine various semiotic systems. As a result, having taken Clark's view into account, it is possible to claim that language, being a system of signs, is inherently multimodal. It does not necessarily mean that people make use of all the possible modalities/semiotic systems in each instance of interaction: the distribution of expressive means varies depending on a given communicative situation, which is why it would probably be more correct to treat language as variably multimodal. Considering the complexity of human social interactions, however, it is hard to find a situation where different modalities and different modes within one modality do not interact, at least to some extent, in most communicative situations, whether face-to-face or without any visual contact with their interlocutors, people employ a range of expressive resources beyond language in its traditional interpretation. Such resources are derived from both vocal (e.g., prosodic elements such as intonation, stress, and rhythm) and bodily modalities (e.g., gestures, facial expressions, body orientation and posture).

*In other words, the multimodal expressivity of language is the key to its adaptivity and interpretability*. When it comes to language comprehension, people tend to combine information received from different modalities and integrate this information into the process of language comprehension, while in the course of language production, we

[1] "Spoken" is used here to differentiate between oral/aural modality as opposed to manual/visual.

frequently rely on the affordances inherent in each modality. Moreover, it is the variety of semiotic resources, together with the contextual information and global knowledge of the world that allows to overcome ambiguity. In other words, gestures, along with other expressive means traditionally treated as non-linguistic elements, ensure communication efficiency, and help decrease processing costs in conditions of structural and semantic ambiguity.

One of the examples that demonstrate our reliance on the simultaneous presence of multiple modalities in order to ensure successful communication is related to the use of emojis in digital communication, although the body of research on the role of emojis and similar digital behaviors has started to grow relevantly recently, their significant effects on the quality of communication has been unquestionable. They can make explicit the tone of the message and other interpersonal functions otherwise lost in the absence of face-to-face interaction (Kaye et al., 2017). A similar thing can be said about the use of emojis representing various hand shapes. This particular example will appear again further in the thesis when discussing the grammaticalization of gestures.

In other words, meaning is not transmitted solely by language, but more often by a vast number of resources including gesture. Elaborating on this in terms of cognitive approach to language, gestures play, among other things, an important role in cognitive processes such as communication, encoding and recall. They allow speakers to express thoughts that do not fit into the categorical system of a given language, they facilitate comprehension by providing disambiguation cues, and they enhance communication between the speakers. Moreover, as it will be repeatedly shown in the following chapters, gestures also can reflect the properties of humans' conceptual structure, driven, in turn, by the nature of our (embodied) cognition. This capacity of gestures is reflected in research by rethinking concepts from cognitive semantics such as image schemas, conceptual metaphors, and metonymy (Bressem, 2021; Cienki, 2013; Mittelberg & Waugh, 2009; Zlatev, 2014, inter alia). Moreover, gestures have the potential to be analyzed in terms of their grammatical and semantic properties in a way similar to that of words and complex phrases and, thus, can be structurally and/or functionally integrated into language system.

Having said that, it is important to note that albeit acknowledging the possibility of structural and/or functional integration of gestures into speech, the current work refrains from claiming that the use of gestures (as well as signs in sign language) is guided by the same set of underlying principles as the use of words and constructions in a spoken language; rather, borrowing the conceptual and terminological apparatus from the Jackendoff's framework of the Parallel Architecture (Jackendoff, 2004), we entertain an idea (in line the claim made by, for instance, Fricke, 2014) that gestures in a broad sense constitute an autonomous system which frequently interfaces with spoken language reserving at the same time an ability to bypass it while still functioning for the purposes of communication.

Hence, the first premise of this work is that speech and gesture, despite being tightly intertwined, are, to an extent, different in their expressive means and structural properties. The observed overlap between these two modes of expression may be coincidental, primarily due to the shared conceptual knowledge, human cognitive abilities, and the dominant role of spoken language among the hearing population. As will be seen throughout the thesis, the research up until now looked at co-speech gestures through the prism of our knowledge about the system of spoken language; however, it may be useful to focus on differences between the two modes of expression in order to answer the questions of what gestures do that speech does not or what the properties of gesture are. In other words, the significant overlap between the structural and semantic properties of gesture and spoken (and sign) language may deviate attention from the gesture proper. These questions can be more important from the evolutionary and cognitive points of view, as they can shed light on the nature of human communication beyond words. Given the theoretical nature of the work, the main attempt will be to present existing knowledge regarding the use of co-speech gestures and their relation to speech using *a pragmatic approach*, as researchers on co-speech gestures and their role in communication tend to look at gestures as a linguistic (sub)-system guided by the same or very similar organizational and structural properties as spoken language. *So, this knowledge about gestures will be used as a starting point, but with the idea of pointing out inconsistencies and ambiguities to highlight the potential need to move beyond such descriptions.* Hence, in the current work, the elements of the proposed framework will be separated in the following way. The properties of perceptual cognition giving rise to conceptual knowledge about the world will be discussed separately and treated as one of the primary reasons for the presence of similarities on different levels between spoken language and gesture, while language-proper notions will be covered in separate chapters.

The second important assumption in this work regarding the role of gestures in communication is derived from the perspective of language evolution. According to a recent view, language has largely developed as a bimodal system, where the expansion of early hominin communication based on the rich gestural repertoire was closely followed by the growing range and complexity of vocalizations. In other words, the complex structure of spoken language is seen to have evolved as a result of a progression of increasingly syntactically rich protolanguages, together with the transition from predominantly gestural to predominantly vocal forms of communication (Planer & Sterelny, 2021). If assuming that language originated from the sequencing and subsequent integration of two modalities (vocal and bodily), we also consider both the systems of sign and spoken languages, it is possible to make two auxiliary hypotheses. Firstly, the features of both modalities may be observed in both independent systems of sign and spoken languages. Naturally, this is true for vocal modality, but it also stands for manual communication: the research has shown that the use of spoken language elements, such as mouthings, is pervasive among sign language users. It is likely to have something to do with the inherent breath and manual movement coordination or, in other words, speech-gesture biomechanics (Pouw & Fuchs, 2022). This idea is not treated in this work as a self-standing hypothesis but as one of the arguments for why co-speech gestures should not be overlooked in linguistic research. Secondly, and most importantly, if we look at sign and spoken languages at their current stage of development as a result of the evolutionary expansion of proto-languages, which relied on bodily modality as an initially richer source of meaning construction, one may argue that it would be wrong to dismiss speech-entangled gestures from the analysis as an unlimited set of unordered and unmotivated movements; rather, we should treat gesture as a generative system in its own right, following similar principles of organization and form-meaning relations as fully-fledged linguistic systems. Having said that, we can look at the system of gestures as a precursor of modern sign languages, which has also left its traces in spoken language. At the same time, it would be wrong to assume that, given the primacy of bodily modality from the adopted evolutionary perspective, gestures can be treated as evolutionary fossils rather than an integral part of language at its current stage of development.

Considering the abovementioned premises, current work focuses on answering the following questions:

**RQ1: What lies in the core of the significant overlap between gesture, the system of sign language, and spoken language?**

Our knowledge about the world largely determines the way we conceptualize reality. Through daily interactions with various objects, humans learn the possible physical properties and relations between those objects. Moreover, through the experience of different events, such as physical movement or mental state, we become aware of how certain actions and physical changes affect our bodies and the surrounding reality. These and similar experiences contribute to building internalized knowledge about the world, which is employed for a range of behaviors, including instrumental actions and, naturally, communication and language. It is argued in Chapter 2 (Spatial Structure and Conceptual Structure) that this knowledge is stored as a system of perceptual symbols, following the seminal work by Barsalou (1999), as opposed to a widely accepted view of the amodal symbol systems. Having outlined Barsalou's theory, we will present some compelling evidence from the recent research on the relation between vision and cognition. Hence, the argument developed in this work is that the conceptual system is an autonomous system which can interface with both the linguistic systems and gesture (whether the latter share the levels and to what extent is discussed as a part of RQ3) and that spoken language and gesture share computational resources (the idea developed as a part of RQ2), ensuring their mutual penetrability.

**RQ2: How can gesture be integrated into the syntactic and semantic structure of verbal utterances?**

Using both the contributions from theoretical studies and results from existing research obtained in the past thirty years, the thesis explores how the concepts of cognitive linguistic theory spanning morphology, semantics, and syntax can be applied to the analysis of speech-entangled gesture. The work focuses on each of the linguistic levels separately, drawing parallels between the underlying structural principles of spoken languages and their manifestation in co-speech gestures. It is shown that the vast majority of theories were developed to describe how principles of organization in spoken language can be extended to explain the functioning of gestures used in combination with speech. In particular, Chapter 3 (Morphology) shows that certain parameters of gesture form can establish stable form-meaning connections and become the core of several gestures forming a gesture family. As a result, it is claimed that co-speech gestures possess rudimentary morphology, which manifests itself in the existence of a number of form parameters that influence the meaning of a given gesture within the shared semantic field. Chapter 4 (Semantics) discusses how gestures show consistent behavior with respect to similarity, contiguity, and schematization. Importantly, it supports the idea that gestures must have an internal semantic structure and conceptual hierarchy of their own, which is, in turn, to an extent, guided by our embodied experiences with the physical world. Lastly, Chapter 5 (Syntax) explores the grammar of gesture, looking at two processes which reflect different degrees of freedom of gestures from spoken language: code integration dealing with ways of syntactic integration of gestures into speech, and code manifestation reflecting more advanced levels of gesture autonomy. The discussion in this chapter also focuses on the tension between linear and hierarchical structures in gesture and provides a closer view of phenomena such as negation, aspect, and continuity.

However, the analysis of both theoretical sources and experimental findings leads to identifying the major caveat of the modern approaches to co-speech gesture, which stems from the attempt to assign the features of spoken language to gesture as a way to find evidence for the presence of similar structures in both modalities. As a result, we look at gestures from the perspective of a complex system of spoken language shaped and refined in the course of evolution. However, this approach is not comparable with the existing evolutionary view that states that language emerged in a predominantly gestural form, setting the course for predominantly spoken communication only later. Thus, we can consider spoken language at its current stage as the result of humans' cognitive development over thousands of years, which by no means suggests that the system of language in its original gestural form is comparable with the system of spoken language nowadays. Moreover, it is very likely that a rich system of spoken language highly influenced co-speech gestures in the form they exist today. Hence, instead of analyzing similarities and differences in gesture within a spoken language paradigm, the current work proposes that it may be useful to seek the features of gestures absent in spoken language.

**RQ3: How can the unity of gesture and speech be explained using the multimodal Parallel Architecture?**

This work attempts to answer how the unity of gesture and speech components can be explained following the framework of (multimodal) Parallel Architecture (Cohn, 2016; Cohn & Schilperoord, 2022; Jackendoff, 1985, 2004). Here, apart from the notions of morphology, semantics, and syntax outlined above, an important role is assigned to Spatial Structure as direct instances of our embodied experiences in a physical world. In particular, the present work focuses on visual-spatial processing as one of the key sources which provide humans with information about the surrounding reality, as explained in Chapter 2 (Spatial Structure and Conceptual Structure). It is important to note that the model proposed in Chapter 6 (Parallel Architecture) is not much more than a framework used to show how various non-verbal behaviors and semiotic resources can be seamlessly integrated into a single unified framework. The Parallel Architecture framework is treated as flexible and adaptive enough to account for the multimodal nature of language; however, it does not bear any explanatory power. While it shows that Spatial and Conceptual Structures are likely to be shared among different modalities interfacing with other linguistic structures, it does not give definite answers to how exactly all these structures manifest themselves on various levels of language and linguistic principles these structures are grounded in. In other words, while the system of co-speech gesture is tightly intertwined with the system of spoken language, we cannot definitely say which of these elements constitute gesture proper and which emerged as a result of centuries-long co-existence of gesture and language made possible due to the shared conceptual system.

The last Chapter 7 (Discussion and conclusion) of the thesis is dedicated to a more extensive description of the problems dominating the research in gesture, walking through the terminological and conceptual ground lain in the main part of the work to identify gaps in the present-day research and outline the favorable course of investigation in the future.

## 1.1 Why gesture?

### 1.1.1 Background

The main reason why the research on gesture and other non-verbal behaviors emerging in the process of communication had not fallen into the focus of linguistics until the 1990s stems from the view on language established since the earliest principled attempts in the late nineteenth century to tap into the nature and organization of human language and communication. Traditionally, linguists focused on defining an object of linguistics as narrowly and accurately as possible, thus restricting the scope of phenomena they aimed to investigate and analyzing language independently of its speakers. In particular, special emphasis has been placed on the arbitrariness of language as a feature unique to humans, excluding other forms of communication as not "truly" linguistic. This idea traces back to Ferdinand de Saussure, who characterized language as a system of arbitrary signs established for each language as a form of convention between speakers. As a result, other elements of communication were excluded from the linguistic analysis as too idiosyncratic and lacking systematic organization. Even with the emergence of pragmatics as a separate subfield, the interests of the researchers stayed within verbal aspects of communication.

The gradual shift toward acknowledging the multimodal nature of language started with the growth of knowledge of sign languages in the 1960s. Naturally, the approach to language at the time considered sign languages no more than mere pantomime lacking any language-like features and overall unordered. However, it began to change with the publication of the monograph by William Stokoe called "Sign language structure: an outline of visual communication systems of the American deaf" (Stokoe, 1960). Stokoe showed that sign languages possess a system comparable to spoken language that organizes units, or signs, into consistent patterns that obey certain morphological, semantic, and syntactic rules. Using a structuralist approach in his analysis, Stokoe demonstrated how Hockett's design features of language, such as semanticity, discreteness, duality of patterning, etc., manifest themselves in what is now known as American sign language (ASL). This meant that the generally accepted view on what constitutes a 'truly linguistic' subject had to undergo considerable changes. As a result, the idea of the heterogeneous nature of human communication (always present but never taken seriously) gained much closer attention. One of the first terms re-examined was the notion of the linearity of language structure. Bodily modality, unlike vocal production, which is restricted exclusively to the temporal and auditory dimensions but can exploit a much greater range of resources: time, three dimensions of space together with the greater affordances of hands, eye gaze, head, and torso, all allowing for the simultaneous encoding of several types of information. However, these features of bodily modality were overlooked with application to spoken language still being dismissed as involuntary and idiosyncratic actions or, at best, components of meta-communication until the 1990s, which were marked by the works of two psychologists, Adam Kendon and David McNeil, who built a major part of their research primarily around the use of gestures in conjunction with speech. They demonstrated that the forelimb movements (together with gaze direction, head movements, and body orientation) share their role in packaging the relevant information with spoken language due to their recognizable semantic characteristics and transparency of their meaning that can be understood albeit in very broad terms.

The Growth Point theory is one of the first and the most powerful accounts of co-speech gesture proposed by David McNeil in 1992 (McNeill, 1992). He was among the first to notice that gesture and speech are "systematically organized in relation to one another" (McNeill

& Duncan, 2000, p. 142). According to the theory, language and gesture are co-dependent and arise from a single conceptual system. This also means that the information from both speech and gesture is integrated into a single mental representation.

Although the Growth point theory proposed by McNeill is still used in the current approaches to gesture (e.g. Information Packaging hypothesis of gesture production (Kita, 2000)), some of his main assumptions proved wrong. On the one hand, McNeill, who was one of the pioneers of gesture research in linguistics, saw speech and gesture as elements of two separate, albeit connected, systems, where gestures are imagistic, holistic, and synthetic as opposed to language, which is characterized by its arbitrariness, analytic nature and linearity. According to McNeill, these systems reflect two fundamentally different types of thought recruited for the same goal of ensuring successful communication. On the other hand, he argued that the system of gesture is incomparable to the system of language due to its non-combinatoriality and the absence of hierarchy. In turn, the ideas put forward by Adam Kendon reflect the current stance on the topic more closely. He claimed that gestures form hierarchical structures and can be decomposed into smaller meaningful units. As will be shown in the next chapters and, in particular, in the chapters dedicated to morphology (Chapter 3) and syntactic properties of gestures (Chapter 5), gestures indeed show capacities for generating complex structures both on a level of individual gestures, gestural compounds, and even the whole utterances. In other words, gestures produced together with speech exhibit similar properties to the sign in sign language, which follow their own phonological, morphological, and syntactic principles. Moreover, Kendon's view on evolution as a simultaneous development of gesture and speech turned out to be contingent on the current views advocating the multimodal roots of language.

### 1.1.2 Contemporary view on gesture production and perception

The contemporary approach to gesture studies which uses for the most part the concepts from cognitive linguistics will be presented in the coming chapters; however, there is an important branch of co-speech gesture research which, albeit not covered in detail in the current work, is fundamental for our understanding of how gestures function in communication. Although co-speech gesture research has been developing for several decades, the question of the autonomy of co-speech gestures remains open (the term "autonomy" will be discussed in detail in the section 1.2 Problems addressed in the work), since their form-meaning relations can often be established only if analyzed together with other modes of expression and modalities, most often speech. Moreover, the roles of such gestures can be defined only through their relations with the meaning of an utterance as a whole. To this date, researchers have not fully agreed on to what extent and how gesture and speech interact. For instance, many insights also come from the research on the syntactic potential of gesture.

The research by Schlenker and Chemla (Schlenker & Chemla, 2018) has shown that co-speech gestures are not only frequently used to replace verbs in a sentence, but also display properties of the American Sign Language (ASL) 'agreement verbs'. A more detailed account of the syntactic integration of gesture into spoken language will be discussed in detail in Chapter 5 (Syntax). Nevertheless, as a result of the observations like the one mentioned above, a number of approaches were developed over the last thirty years concerning both perception and production in an attempt to explain what motivates multimodal expressivity of language.

Currently, two of the leading theories describing interrelations between speech and gesture in relation to the stages of language production are Lexical Retrieval Hypothesis (LR, henceforth) and the Information Packaging Hypothesis (IP, henceforth). According to the LR (Rauscher et al., 1996), the fact that gesture often occurs cataphorically preceding verbal utterances, as well as the evidence that in certain conditions, restriction of bodily/manual movements leads to verbal disfluencies (e.g. Cravotta et al., 2021; Morsella & Krauss, 2004), suggest that co-speech gestures are involved in the stage of language formulation (lexicalization) by facilitating lexical retrieval. The supporters of the IP (Kita, 2000), in keeping with the Growth Point theory proposed by McNeill (1992, 2005), argue that the interaction between speech and gesture starts at an earlier stage of language production. In particular, they claim that gestures play a significant role in information conceptualization. Although the evidence for these theories remains conflicting, both LR and IP hypotheses highlight the importance of co-speech gesture and note systematic temporal co-occurrences between gesture and speech. While the key trigger is temporal co-placement of speech and co-speech gesture, the patterns of gesture-speech alignment and the distribution of dominance might vary.

From the perception point of view, the question of gesture-speech interplay is also related to the problem of sign filtration: all the co-speech gestures always have the potential to be interpreted in terms of their form-meaning relationships with the overall context of an utterance and "are therefore available for inclusion in a unified utterance interpretation" (Enfield, 2013, p. 699); however, they are not always analyzed by the receiver. In other words, such gestural signs used in isolation can be analyzed as tokens (not tokens of types) or so-called 'singularities' (Kockelman, 2005) that become signs only being used in a certain context. Hence, "the problem of comprehending gesture meaning is taken to be one of interpretation (from token form to token informative intention)" (Enfield, 2013, p. 697). Recent research has shown that humans do not tend to privilege linguistic information over other types of cues with very similar patterns of syntactic and semantic processing of both speech and gesture  (Hagoort & Van Berkum, 2007; Özyürek et al., 2007; Peeters et al., 2017; Wolf et al., 2017 inter alia). These findings suggest that when interpreting speech, the human brain makes use of multiple sources of information present in the context, including gestures and facial expressions. Moreover, the fact that listeners can benefit from the information conveyed by gestures may also affect the way people gesture while speaking. For instance, research has shown that the types of gestures people use depend on visibility conditions (Bavelas et al., 2008). In other words, speakers adapt their bodily expressions to the anticipated needs of their listeners.

As can be seen, modern approaches to gesture are largely concerned with the processes of perception and production and, namely, with the role co-speech gestures play in making meaning and drawing inferences. What is important here is that a multimodal view on language and communication is coming to the fore, and more attention is being paid to what traditionally was considered non-linguistic behavior. This thesis, however, focuses more on the theoretical assumptions regarding the place of co-speech gestures in relation to the theory of language rather than the process of linguistic communication.

### 1.1.3 Evolutionary perspective

The developing research on co-speech gestures as well as sign languages, opened up a new chapter in the field of language evolution. In this section, two competing views on the evolution of language will be discussed, which will, in turn, lead to the examination of the concept of arbitrariness as one of the design features of language. Moreover, an approach based on the biomechanical properties of the human body will be briefly mentioned to show new methodologies being explored to offer new evidence for the role of gesture in language evolution.

### A) Gesture first or gesture-speech together?

From a gesture-led point of view (Corballis, 2017; Tomasello, 2010), the emergence of language structure is related to the expansion of early hominin communication based on the rich gestural repertoire, considering the existence of a significant overlap in gesture production among humans and non-human primates with relatively limited range of vocalizations. For example, children of 1-2 years of age share 89% of their communicative gestures with chimpanzees and up to 96% with African apes (Kersken et al., 2019). It has also been shown that there are similarities in the meaning of gestures used by various species, which are greater than expected to happen by chance (Graham et al., 2018).

However, although gestures indeed are characterized by bigger repertoires and greater flexibility in comparison to vocalizations, it was shown by a number of comparative studies (e.g. Liebal, 2014; Taglialatela et al., 2011) that human and non-human primate communication is inherently multimodal. The researchers focusing on the comparison between humans and non-human primates also observe a number of similarities related to the features critical for successful interaction: intentionality, referentiality, iconic character and combinatorial nature of the produced signs, turn-taking, neural control, and ontogenetic plasticity. Moreover, although it is too early to draw any definite conclusions regarding to what extent non-human primates possess these characteristics, there is undoubtedly a substantial amount of evidence that the overlaps in cognitive characteristics between humans and non-human primates can be manifested in in either one of the modalities (vocal or bodily) or both modalities simultaneously (see the extensive review in Fröhlich et al., 2019).

It is important to note that despite these similarities, some major changes in the use of gestures (and other bodily behaviors) and vocalization must have occurred at a certain stage in the evolutionary development from great ape communication to the language we know today. This transition could have occurred at the stage of proto-language use by the earliest hominins. J. Planer and K. Sterelny (P&S) hypothesize "an initial transition from non-composite to composite communication, and then from regular composite sign use to simple syntax using linear order" (Planer & Sterelny, 2021) or, in other words, progression of increasingly syntactically rich proto-languages together with the transition from predominantly gestural to predominantly vocal forms of language. Moreover, P&S contend that much of the hierarchical complexity of spoken language may have evolved as a result of the development of preexisting action/planning systems, an argument important for the Chapter 2 (Spatial Structure and Conceptual Structure) of this thesis.

P&S (2021) argue that there are two main evolutionary changes that led to the emergence of the systems of communication: the transition to bipedalism, which freed the upper body and allowed for better motor control in the upper limbs, increasing complexity of the behaviors which called for the importance of their transmission from one individual to another. P&S see importance in the links between language and cultural, economic, and technological advancements in humans and early hominins. In particular, one of the major advances that, according to the researchers, impacted both the organization of life and biological changes was gaining control of fire, which was associated with a number of cooperation and coordination problems. The presence of a source of light available even when the sun was out, as well as the complexity of the process of starting and maintaining the fire in itself, enhanced the demands on social life, which was one of the crucial steps when the transition from predominantly gestural to predominantly vocal form of communication occurred. Moreover, the presence of fire affected the feeding habits of early humans: the consumption of processed foods compared to raw foods was likely to take less time and, more importantly, influenced the changes in the speech anatomy, making it more suitable for speech production over time. This allows for an influential assumption: the evidence suggests that the first control use of fire dates back 400,000 years; however, the need for some type of communication between individuals must have existed even before. This leads to the conclusion that more elaborate gestural forms of communication preceded the advancement of the vocal ones.

B) The issue of arbitrariness

As mentioned above, most research in linguistics also highlighted the importance of arbitrariness and the symbolic nature of language, which elevated its status compared to other systems of signs. The reason is that arbitrary signs have been seen as the features of later, more advanced developmental stages in language evolution due to higher processing costs associated with their production. Icons and indexes, in turn, have been treated as less cognitively demanding and, thus, more primitive means of conveying meaning. In other words, the linear view on the development of the linguistic system from idiosyncratic signs to highly conventionalized form-meaning pairings is the main driving force of language. However, although this view is not entirely wrong, it provides a simplified perspective on language development, overlooking a number of important facts. Firstly, the nature of indexes and icons is much more versatile than was suggested before. Both indexes and, especially, iconic signs can differ in the degrees of conventionality and schematicity, as well as according to the importance of socially shared conventions and cultural coadaptation. As a result, most indexical and iconic signs represent hybrid versions of signs with both abstract and arbitrary features. Secondly, the question of cognitive demand also remains open. Iconic representations seem to be nearly as demanding as arbitrary signs and symbols (although the concept of arbitrariness as a direct opponent of iconicity is nowadays also criticized as being seen as a result of the human capacity to establish conventions (e.g. Winter, 2021)) in the level of awareness required for inferring and interpreting other speaker's intentions together with the ability to anticipate communicative moves and the knowledge of the cultural background (Planer & Sterelny, 2021). The absence of neurological effects of iconicity versus arbitrariness hass been also shown in the study which focuseed on the ERP effects (N400) of iconicity during lexical access for signs (Emmorey et al., 2020). This study showed no significant difference between the neural activity when processing iconic versus non-iconic signs when also controlling for frequency and concreteness.

In turn, a number of developmental studies following the changes in language abilities in children of various ages support this view. For example, the results show that children start producing iconic gestures six months after uttering their first verbs, which suggests that the former are associated with greater conceptual difficulties (Özçalişkan et al., 2014). At the same time, the understanding of iconic gestures is still better in children between 24 and 36 months old than the understanding of iconic vocalizations (Bohn et al., 2019). This suggests that gestures play a significant role in language ontogeny.

C) Speech-gesture biomechanics

Some of the evolutionary arguments related to the role of gestures are grounded in understanding the biomechanics of movement. Firstly, from the evolutionary point of view, the shift to bipedalism freed the hands and led to the development of more advanced motor control. Secondly, Wim Pouw and his colleagues argue that the origins of gesture and the following development of sophisticated vocal abilities in humans lie in physical relations between the actions of the body, respiration, and vocalization (Pouw & Fuchs, 2022). According to the researchers, "beat" gestures following the prosodic characteristics of speech might have served as a preadaptive mechanism connecting bipedalism and respiratory-motor coordination, which led to a complex vocal system in humans. The evidence for that is found in pre-verbal infants whose verbal development is preceded by vocal-motor babbling (MacNeilage, 2010).

### 1.1.4 Developmental perspective

The relation between perception and communication is observed in the developmental literature. For example, when it comes to the stages of language acquisition in infants, it is pointed out that the production of gestures, especially pointing and iconic ones, often precedes the holophrastic stage. Moreover, researchers argue that an important intermediate stage that signals an approaching transition from the holophrastic stage to telegraphic speech is the production of multimodal combinations consisting of a word plus gesture (Cochet & Vauclair, 2010; Fasolo & D'Odorico, 2012; Iverson & Goldin-Meadow, 2005 inter alia).

If we accept the view that perceptual mechanisms lie in the core of our conceptualization of the world, certain patterns of children's language development come around as natural. Firstly, consider the semantic constraints such as noun bias, whole-object bias, and shape salience, and taxonomic constraint which allow children to establish a link between linguistic labels and concepts. All these biases are clearly of a perceptual nature. However, there is more.

In one of the early research on gesture (Stephens & Tuite, 1983), it was argued that iconic gestures can be subdivided into two types (iconix$^1$ and iconix$^2$) depending on the proximity of the given gesture to its source. Nowadays, this classification has become largely obsolete; however, it comes in handy when looking at the gestural developmental trajectory in hearing babies. In particular, it has been observed that the use of iconic gestures follows a certain order: first, children acquire iconix$^1$ only later followed by iconix$^2$. The reason is that infants' iconic gestures tend to be more detailed than similar gestures produced by adults. For instance, as McNeill mentions in his 1985 influential paper, "children prefer to make gestures showing running, not with their hands, but with their feet, and iconic gestures incorporating the head and legs are far more common with children that with adults" (McNeill, 1985, p. 364). Thus, children's first gestures as well as first words reflect "interactive routines between the baby and objects and people" (McNeill, 1985, p. 363).

Here, it may also be useful to bring about a distinction between N-learning vs C-learning (Chater & Christiansen, 2010; Christiansen & Chater, 2022). "N-learning" (N stands for "natural") entails the process of learning to navigate natural world, such as, for example, the knowledge about the persistence of objects, gravity, etc. “C-learning” (C stands for “cultural”), in turn, refers to the knowledge that is transferred through social interactions with other people. While Christiansen and Chater (2022) make the claim that language learning is largely a product of C-learning, it is difficult to agree entirely. While cultural transmission is undoubtedly crucial for the acquisition of language, the physical experiences in the world are no less important. During the first years of their life, even preceding the production of first words, children develop a profound understanding of the physical world (it is also argued that some of this knowledge may be innate, such as basic mathematical reasoning (e.g. Schwartz, 1995)). By selectively focusing attention on aspects of experiences, infants create complex real-world knowledge (Bedny et al., 2008; Jones & Smith, 1993), which is actively employed when encountering culturally and socially shared linguistic concepts. Hence, human communicative ability is likely based on the continuous interplay between N-learning and C-learning.

What implications does all of it have for the use of gesture and bodily modality more broadly? It is possible to argue that N-learning facilitates not only the use of gesture as has been seen in the example of the use of iconix1. It also implies that gestures can be naturally extended to their use in communication and can potentially bypass language. This argument will be discussed again further in the work (esp. Chapter 6. Multimodal Parallel Architecture)

## 1.2 Problems addressed in the work

This work aims to address the following problems.

Firstly, in Chapter 2 (Conceptual Structure and Spatial Structure) we will focus on the two levels which are argued to lie at the basis of our ability to use language (albeit not being the only elements of human language faculty): the biological mechanisms of perception, in particular, vision and proprioception, and the cognitive ability to build mental representations based on the perceived input (both non-linguistic and linguistic). Using the term offered by Jackendoff in his PA, these two levels can be called Spatial Structure (SpS) and Conceptual Structure (CS) respectively (more on SpS and CS in Chapter 6 Multimodal Parallel Architecture). These two levels can be seen as the properties of cognition not unique to language, but yet crucial for language use. In turn, the following chapters will focus on mechanisms more closely associated with language systems (i.e. morphology, syntax, and semantics). Importantly, this thesis does not provide a solution to the “translation” problem. In other words, there will be no proposal regarding how language emerges from perceptually grounded conceptualizations, but rather an overview of how these mental models function in relation to the components of language. Hence, we only argue that there is a link and give compelling indirect evidence for its presence, but what this link entails goes beyond the scope of this research. Thus, the discussion of the cognitive mechanisms of perception will be systematically completed by the ideas of how they may be conceptualized and subsequently recruited in human communication and, in particular, in the use of gesture.

Turning to gesture as a component of spoken communication, the first issue covered in Chapter 3 (Morphology) and section 4.1 (Gesture classifications) of Chapter 4 (Semantics) will be the heterogeneous nature of gestures, which makes them difficult to analyze and systematize. Given the complexity of forms and diversity of communicative functions, linguists face a complex task of developing a reliable terminological apparatus that would allow for consistent analysis of such behaviors. Most representation gestures are non-conventional and depend on the gradience of relations between form and meaning: one form may refer to a number of unrelated meanings. Moreover, given the absence of clear boundaries between different types of bodily movements, it isn't easy to separate linguistic from non-linguistic behaviors. As has been shown before, the basic idea that arbitrariness is one of the most important and curious features of natural language, which places greater cognitive demands on the human brain than icons and indexes, does not receive convincing empirical evidence. On the contrary, research shows that although it seems true that symbolic signs emerge later in language evolution, language largely depends on the use of icons. One more by-product of this lack of basic methodological unanimity is that comparing co-speech gestures and signs in sign languages becomes difficult, if at all possible. By creating a universal approach to the analysis of gesture in a broad sense, the principled comparison between different expressive forms in bodily modality may also become more feasible. In this work, a number of approaches to gesture classification are described and analyzed from the point of view of their potential for interdisciplinary research.

The third issue is related to the degree of autonomy of the system of gestures, which is tackled in particular in Chapter 4 (Semantics) and Chapter 5 (Syntax). Naturally, one cannot expect to find the presence of a richly developed grammatical and semantic organization in co-speech gesture, having, among other things, accepted the evolutionary perspective where gestures are precursors of sign and spoken languages. In order to deal with these issues, it is important, first and foremost, to define the term "autonomy". By using this notion here, we do not argue that speech-entangled gestures can function independently of speech (thus opposing autonomy to independence). However, the argument is that the use of co-speech gesture is guided by certain sets of rules that overlap with those of both spoken and sign languages to various degrees but also display their unique properties. In such case, gestures can be treated as an autonomous system with possesses what one might call a rudimentary linguistic structure with its own principles and regularities. This allows researchers to treat co-speech gestures similar to words and word constructions in spoken (and signed) language and, thus, allows for using metalanguage and terms originally invented to refer to linguistic concepts. Nevertheless, this approach has three major limitations. Firstly, it is still difficult to define the strict boundary between the highly ordered system of signs in sign languages and the system of co-speech gestures. Secondly, using the theory developed to analyze spoken language may divert researchers' attention from the unique properties of gestures. Lastly and most importantly, the abovementioned approach is useful as long as we admit that the way we use gestures in conjunction with speech is likely to be highly influenced by the system of spoken language and the current stage of humans' cognitive development. Using only this approach, we cannot go beyond the descriptive account of the current role gestures play in communication and draw any conclusion regarding what constitutes gesture proper and how gestures emerged and evolved in the course of language evolution. In order to do the latter, we have to consider evidence from other sources. For instance, because of the recent achievements in cognitive science, we know that the components of meaning do not arise independently from a one-to-one mapping between the signs (e.g., words) and the physical concepts in the world but rather due to certain structure of human brain and its ability to conceptualize the world through various bodily experiences. Thus, it is possible to conclude that representational gestures might have emerged as a natural first step in referring to and communicating about these embodied experiences before moving towards a highly conventionalized and abstracted system of arbitrary signs.

Lastly, in Chapter 6 (Multimodal Parallel Architecture), the thesis addresses the issue of developing a reliable framework for the research on gesture (along with other modalities). One of the problems linguists interested in what was traditionally considered non-linguistic behavior face is the inability of the majority of existing theories of language to seamlessly accommodate the variety of resources employed in "doing language". One of the theories which closely reflect this view on communication is the most recent interpretation of Clark's theory on methods of signaling described earlier (Hodge & Ferrara, 2022; Hoffmann, 2021). However, while acknowledging the variety of semiotic behaviors deployed in the process of communication, it does not attempt to explain how these modes interact and how they may correspond to the linguistic structures on the levels of morphology, syntax, semantics, etc. Hence, in the current work, it is argued that one of the most promising accounts that can explain how the systems of speech and gesture interact across a continuum is Jackendoff's framework of the Parallel Architecture (henceforth PA). Following the ideas of the PA, a more nuanced and integrative approach to language can be provided, where multimodal expressions are defined in terms of their component structures and interfaces between them. In particular, the research will focus on how spoken and visual modalities interface with perception and not language-specific cognitive principles, and the mode of gesture can accommodate various linguistic components, such as morphology, syntax, and semantics and show how grammar and meaning interact across different modalities.

The discussion of strengths and weaknesses of co-speech gesture research, together with some proposals regarding the future of the field and the main conclusion, are presented in Chapter 7 (Discussion and Conclusion).

# 2 Spatial Structure and Conceptual Structure

As has already been pointed out, the general heuristic of semiotic theory/pragmatic unity heuristic states that several signs produced together should be seen as a whole and interpreted as a single (decomposable) composite unit. In line with this view, speech-with-gesture composites interact with each other and this way, contribute to meaning making. This makes it possible to talk about the continuity between various modes of expression in the process of meaning construal and comprehension. However, solely analyzing the semantics of these forms of communication is not sufficient to understand how meaning arises. Thus, it is important to establish the link between communication and other cognitive processes, such as perception, attention, and memory, as it is widely accepted that "our cognitive structures are formed in a constant interplay between our minds and the external world" (Gärdenfors, 2014, p. 10). As Gärdenfors argues, the main goal of semantic research should consist of solving the "translation problem". i.e., developing a theory about how the meanings of linguistic elements can be linked to our perceptual systems and physical actions. Importantly, the link between the physical world and the linguistic ways of expression is not direct. Instead, it is mediated by the conceptualizations of the real world that speakers of a particular language develop and associate with linguistic units.

Thus, we are talking about the interplay between perception, cognition, and language; the connections discussed are not inclusive, as the main focus is on the multimodality of communication manifested in our use of co-speech gestures. In particular, the argument is that the considerable weight is carried by the system of vision (coupled with our physical experiences with these external objects) as a driving force of mental representations that, in turn, allow for linguistic mapping. In other words, language emerged for communication, but not all the components of our ability to communicate using language are language-specific; rather, language, as well as various actions and activities we engage in are determined by our knowledge of the surrounding physical world. This knowledge emerges through our direct interactions with objects/other people/the world. As a result, the physical properties of the objects we interact with/situations of use (and our predictions that are possible based on this information), our personal experiences of physical movements and mental states as reactions to the surrounding reality, memories about objects/people/events constitute this knowledge. In this chapter, the theory of Perceptual Symbol Systems (PSS) by W. Barsalou (Barsalou, 1999) will be used as a main framework to establish the link between perception, cognition, and language.

## 2.1 Introduction to the theory of Perceptual Symbol Systems

The main theory entertained in this chapter is Lawrence W. Barsalou's theory of Perceptual Symbol Systems (PSS), proposed in the late 20th century (Barsalou, 1999). Barsalou argues against the widely supported view of amodal symbol systems, according to which perceptual states are transduced into completely new representations with only arbitrary links between them, and claims that this and similar theories should be viewed "with caution and scepticism" (Barsalou, 1999, p. 580), because not only are they not falsifiable, but also lack a satisfactory explanation of how the process of mapping perceptual states onto amodal symbols occurs. The researcher, in turn, suggests that a more plausible theory would be the one in which the perceptual states obtained through the sensory-motor systems generate symbols whose internal structure is "modal and (…) analogically related to the perceptual states that produce them" (Barsalou, 1999, p. 578). Importantly, such symbols are not just holistic recordings of the perceptual states but rather schematic[2] and compositional[3] representations. This feature of perceptual symbols is attributed to one of the core mechanisms, - selective attention, which focuses on specific (more salient) features of perceived experiences allows for their extraction and storage in a long-term memory: "where selective attention goes, long-term memory follows" (Barsalou, 1999, p. 583). Through the multiple interactions with objects/events and their parts, we build complex networks of features that can be reused separately or several at a time, depending on the task. Barsalou calls the sets of these elements or subregions "frames". The role of the frame is to organize elements within it into categories as well as construct possible *simulations* in the working memory. As a result, the perceptual symbols stored in a long-term memory represent not tokens but types. Such theory allows for a natural explanation of how cognition goes beyond perceptual input: from a recording system to a conceptual system. In summary, the main features of perceptual symbols go as follows: such symbols are schematic, componential, and conditioned by the modality in which they originate. Before discussing the implications of these features for the bodily modality, it is useful to look at each of these properties separately and see how the existing research on perception, cognition, and language can back them up.

Hence, the next section will focus on the description of the abovementioned features of perceptual symbols using visuospatial cognitive principles as a baseline for the discussion regarding how the knowledge about real-world objects and events is represented in the human brain to be used in language and gesture.

## 2.2 Visual-spatial processing

One type of information received from our interactions with the world that is going to be discussed in the present work is information received via visual sensory input, as this perceptual mechanism is "particularly prominent for the construction of experience with the world and the language that organizes that experience" (Mishra & Marmolejo-Ramos, 2010, p. 297). From the physiological perspective, vision acts as the primary source of information that helps people to interpret and categorize the world around them. It guides our interaction with the environment by determining the physical properties of objects, providing spatial cues, and regulating body movements to act upon those objects. Importantly, visual perception also shapes the way humans conceptualize the world (both during learning (and production) and online processing/comprehension), which can be seen in both language and gesture. In other words, "the structure of visual processes (…) constrain[s] semantic representations" (Gärdenfors, 2014, p. 15). Moreover, unlike spoken communication, which greatly relies on auditory processing, bodily modality is a natural extension of humans' ability to visually perceive objects, their spatial characteristics, movement, etc.

[2] schematically stored simple elements of perceptual experiences as opposed to experiences stored holistically.

[3] simple schematic representations that can be productively combined to form complex representations.

The role of spatial-visual perception has been looked at from a linguistic point of view and has given rise to a number of theories. For instance, one of the first such theories is the *Visual World Paradigm*, which makes predictions about how visual and linguistic information is integrated into spoken language comprehension. A set of experiments conducted by a group of researchers in 1995 demonstrated that sentence processing happens incrementally, and the processor relies on different types of information available at the moment. Hence, language comprehension is not mediated merely by linguistic subsystems but also feeds off various kinds of non-linguistic information, such as spatial-visual context, which facilitates the disambiguation process of linguistic input (Tanenhaus et al., 1995). Moreover, our mental projection of the real world is reflected in abstract patterns such as image schemas and more concrete mimetic schemas (see Section 4.2 Schemas). These examples concern the inherent feature of visual processing: refocusing or attention shift from one property of a given scene to another, which also allows for metaphorical and, most importantly, metonymical extension essential for gesture formation (see Section 4.3 Metonymy and metaphor).

However, this theory focuses on the online processing of linguistic input, which is compatible with the idea of pragmatic unity heuristics, especially relevant for the perception of linguistic stimuli, but does not explain the mechanisms that allow for using the knowledge acquired via perception in the process of communication. In this chapter, we will look at a number of theories that attempt to uncover the mechanisms that would make the relations between perception, cognition, memory, and language possible. However, instead of presenting these theories one after another, the focus will be on the features of perceptual-cognitive mechanisms that are frequently discussed (explicitly or implicitly) in the works on the topic and go as follows: schematicity and componentiality, compositionality and productivity, and modality-dependence. Before we proceed with the analysis of each of these three features, there are two things worth mentioning. Firstly, as has been said, in some research, they are not used explicitly as well as the terms themselves can vary from author to author. Secondly, one may notice that the given terms coincide with the ones used in linguistic research; however, *there is no answer to what extent, for instance, compositionality in vision and compositionality in language are guided by the same cognitive principles*. Thus, although the choice of terminology is not coincidental and allows for drawing inferences, as will be seen further in the chapter, there is still a distinction between how these phenomena are treated in relation to perception and in relation to language proper.

### 2.2.1 Schematicity and componentiality

Schematicity and componentiality[4] lie in the core of humans' visual perception as one of the by-products of selective attention and incremental processing of visual stimuli. In this, Barsalou's theory is close to the recent research on language, vision, and cognition carried out in the PAL lab by Alon Hafri and his colleagues. They also claim that certain features are more salient than others, which leads to the so-called "skeletal representations" (Hafri et al., 2023), which exhibit the following properties: discrete constituents, role-filler independence, and abstract content (Hafri et al., 2023, p. 1). The researchers claim that when presented with visual stimuli humans extract certain general features which let them

[4] While these concepts, along with other high-level perceptual mechanisms, are inherently part of semantics, we aim to distinguish here between linguistic and perceptual processes and refrain from claiming that they are exclusively governed by identical cognitive principles.

categorize novel stimuli despite differences in their surface forms, as, for instance, illustrated by figure 2.1, where three separate images present an instance of the same relation of containment.

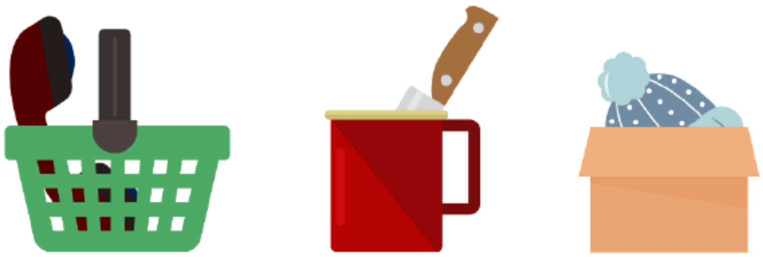


Figure 2.1. Three images sharing the relation of containment (Hafri et al., 2023, p. 2)

it is possible to find the idea of componentiality in earlier works on visual perception. For example, a substantial amount of research in semantics (e.g. Jackendoff, 2004) draws on Marr's idea of 3D model (2010). David Marr was developing visual processing models integrating findings from psychology, artificial intelligence, and neurophysiology. His 3D model of vision aimed to explain how the human brain converts a two-dimensional representation of a single object/scene (in statics) projected on the retina into a three-dimensional output.

Marr presented this model as a three-step process: from a "primal sketch", which is a schematic representation of the key components of the scene/object. In other words, the scene/object is presented as a combination of shape primitives located along the natural axes of an object-centered coordinate system comprising edges, virtual lines, boundaries, etc. At the intermediate stage, depth cues such as shading, orientation, texture, and information about retinal disparity, are used to construct a 2½ -D sketch. It is yet a 3-D model since it is impossible to imagine an object from behind using a single viewpoint. In the last stage, the final 3-D representation is formed using the knowledge of hierarchical organization and modularity in terms of surface and volumetric primitives. Marr also claims that the recognition of the built representation is based on a collection of stored 3D models of stable representations. The process of visual scene encoding happening in the human brain is more complicated than described by the 3-D model. However, there are indeed areas in the brain responsible for processing particular properties of an entity. The model that explains this phenomenon is called the "dual-pathway model of vision".

The successful execution of various visual, auditory, and linguistic behaviors relies on a complex and dynamic interplay of various systems in the human brain. However, it has been argued that the processing of spatial and non-spatial time-invariant information essential for these high-order cognitive processes is distributed between the two functionally specialized processing pathways: the occipitotemporal/ventral pathway, "connecting primary sensory cortices with temporal and prefrontal regions", and occipitoparietal/dorsal pathway, "connecting sensory areas with posterior/inferior parietal and prefrontal regions" (Cloutman, 2013, p. 251). The ventral pathway is primarily responsible for the identification of stimuli (visual, audio) and object recognition, or, in other words, for processing non-spatial information (the "what"), while the dorsal pathway is engaged in processing spatial information (the "where" and the "how"), such as location, the relative positioning of an object in space, distance, motion, etc. Thus, the character of the stimulus is claimed to evoke different responses along these two processing streams. Within the domain of vision, "the ventral stream is engaged in producing a representation of the world for use in subsequent tasks including stimulus identification and memory (vision-for-perception), while the dorsal stream is involved in providing visual guidance for motor action (vision-for-action)" (Cloutman, 2013, p. 252). Despite the claims that such clear distribution of labor between dorsal and ventral pathways is nothing but an artificially created notion, there is a clear evidence from a number of studies using functional magnetic resonance imaging (fMRI) and positron emission tomography (PET) (Anourova et al., 2001; Arnott et al., 2004; Parker et al., 2005, inter alia) that "the brain mechanisms associated with processing spatial (or motor/action) and non-spatial information are associated with distinct and dissociable brain networks in vision, audition, and language" (Cloutman, 2013, p. 253), although there is indeed a strong cross-talk between these dual streams. The dual-steam nature of visual processing was initially discovered in non-human primates (Desimone, 1990; Felleman & Van Essen, 1991) and later extended into the theory of human cognition, which shed some light on the evolution of visual cognition in humans and its role in the development of language. As has been mentioned, that is difficult to isolate these two types of information as it is unlikely that the computations along both of the streams happen independently and in parallel, but rather through a constant interaction between them since our interactions with the surrounding world typically require simultaneous access to both types of the information. As a result, "although the dorsal stream is able to process and execute relatively simple visuomotor processing independently when more complex behavioral responses are required, the integration of semantic information from the ventral stream is essential" (Cloutman, 2013, pp. 253-254).

Marr's proposal, supported by the findings from neuroscientific research, opens the door not only to the understanding of how vision works on the level of perception (bottom-up) but also to the acknowledgement of the top-down processes which rely on long-term memory (recovering 3-D images of objects using directly perceived 2½ -D sketches). The theory also focuses on the incrementality of visual processing, which, as will be discussed further, can be extremely important for the use of gestures. However, Marr's idea does not go far beyond the algorithmic level: while it aims to explain how perception, and vision in particular, works, the 3D model does not resolve the issue of how the sensory input is stored in a long-term memory does not allow for establishing a working theory regarding how the communicative systems can use this perceptual knowledge.

One of the solutions for the abovementioned conceptualization problem can be attributed to another theory by Peter Gärdenfors (2004), who attempts to explicitly explain deep correspondences between perceptual processes and the semantic organization of language. Gärdenfors supports the view that certain basic semantic domains are grounded in our sensorimotor experience. He argues that the core of semantic knowledge lies in the notion of domains, which, in turn, can be largely categorized into two types: physical domains and abstract domains. Physical domains are considered basic and grounded in sensorimotor mechanisms, allowing concepts from this domain to emerge first in a child's linguistic repertoire. At the same time, as long as at least one concept within a domain is acquired, the subsequent acquisition of other concepts of this domain happens at an increased speed. Moreover, throughout language development, domains get separated from holistic representations of these concepts. To illustrate this idea, Gärdenfors uses

Piaget’s (1972)notion of conservation, which has experimentally demonstrated that children until a certain age experience difficulties separating the domains of height and volume when talking about liquids poured into containers of different diameters. Following Jackendoff’s (1985) idea of distinction between the real world and projected world, Gärdenfors emphasizes that the meaning arises not through the direct connection between the language and the reality but rather between the language and the conceptualized meaning spaces created based on our experiences with the world. These meaning spaces arise from our knowledge of domains. While, as has been said, the acquisition of physical domains relies on embodied experiences with the world, learning abstract domains depends on social interactions. For the purposes of the current work, the focus will lie on the physical domain, which consists of the visuospatial domain, force and action domain, and object category space. These domains can be, in turn, decomposed into more narrow ones, such as color, height, width, temperature, force, etc.

The structural properties of these domains can be explained from the perspective of the theory of conceptual semantics (Gärdenfors, 2004, 2014). According to the theory, the human mind organizes information based on primitive quality dimensions that can be explained using geometric/spatial terms. For instance, subclasses of nouns are characterized by particular types of domains and/or the absence of domain (e.g. mass nouns do not have shape domain, etc.). This goes in line with Barsalou’s view, who believes that each simulation stored in a long-term memory gets activated when processing the matching input, which allows to assign new elements into the category with the shared features.  This has interesting implications for linguistic knowledge, such as the theory of language acquisition; however, it can also provide a plausible account for the nature of co-speech gestures.

The main caveat of Gärdenfors’ theory, as well as the theory of PSS, is that they have yet to gain substantial empirical evidence; however, the body of research on this topic has slowly started to build up. Several years ago, a finding was made related to the topological/geometrical structure of the information decoded by the hippocampus in terms of spatial relations and memory (Bellmund et al., 2018; Bush et al., 2015). The most recent study on the geometric nature of human cognition is presented in the article by Sablé-Meyer et al. (2021). The researchers claim that, although it is not unusual for various animal species to possess sophisticated abilities for spatial navigation and production of complex systematic patterns, the capacity to produce “symbolic geometric structures in a combinatorial and productive manner” (Sablé-Meyer et al., 2021, p. 1) remains uniquely human.  Sablé-Meyer et al. also take human linguistic abilities into account, attending to the idea of Hauser et al. (2002), who claim that recursion might be the only uniquely human capacity that accounts for the faculty of language in humans. They hypothesize, following Fitch (2014) and Dehaene et al. (2015), that “recursion is not limited to linguistic communication” and “capacity for recursive syntax and compositional semantics, could underlie many other uniquely human abilities such as music, mathematics or theory of mind” (Sablé-Meyer et al., 2021, p. 3). Another theory that the researchers use to develop their proposal is Leyton’s generative theory of shapes (2004), which suggests that “all shapes are constructed in a bottom-up fashion by a sequence of operations” (Sablé-Meyer et al., 2021, p. 20). Sablé-Meyer et al. propose to differentiate between two strategies required to perform tasks related to the perception of space: purely visual, required for object recognition and shared among several animal species, and abstract or symbolic, presumably available only to humans. Thus, the main argument of their research is that the complexity and compositionality of human thought are achieved by recombining basic patterns derived from our knowledge of space "in order to form complex mental programs" (Sablé-Meyer et al., 2021, p. 20).

In sum, humans do not store each perceived visual stimulus individually in their long-term memory but rather focus selectively on certain features of that stimulus in order to generate mental representations. These representations are schematic and abstract, which allows to assign the object/event to a category or a number of categories depending on which of the features come to the foreground in a given context. The fact that different features are processed separately in the brain is supported by a number of neuroscientific findings. Such compositional view on perception leads to the idea that the conceptualizations we form are also of compositional nature. This is an important implication for the use of representational gestures because they never attempt to represent the objects/events in detail but rather selectively highlight certain properties which can be naturally conveyed via bodily modality.

### 2.2.2 Compositionality and productivity

Productivity means a potential to construct an infinite number of representations from a limited set of symbols with the help of combinatorial and recursive processes. Productivity is seen as one of the design features of language; however, it can be applied to the symbolic representations of objects and events. Using multiple schematic representations organized into categories, humans can simulate, for example, different things and events not experienced directly (as well as things that cannot be experienced or perceived in the real world). Barsalou (1999) imagines the productive process as the opposite of the symbol formation process, during which mental representations are stripped of details and large amounts of information to become schematic. In other words, during productivity, schematic regions are filled with the required information and can undergo necessary changes such as "replacements, transformations, and deletions of existing structure" (Barsalou, 1999, p. 594). Because of the componential nature of perceptual symbols, they can be regrouped and reanalyzed in an infinite number of ways, potentially following certain compositionality principles and constraints.

When talking about how the theory behind perceptual symbols can be extended for use in linguistics, Barsalou follows Langacker's proposal (R. W. Langacker, 2014) according to which grammar corresponds to conceptual structure when it comes to, for instance, the productive aspect of both. In Barsalou's view, humans' ability to construct simulations is tightly linked to our linguistic ability, and there is a reciprocal relationship between them. That is, it is possible both to construct simulations and communicate them via language (e.g., when discussing future events) and to use language in order to construct simulations (e.g., when conveying information about past events experienced by one person but not their interlocutor).

One complex point in the theory is the representation of abstract concepts through metaphorical extensions. Using notions such as *truth*, *falsity*, *negation*, and *anger*, Barsalou contends that all abstract concepts can be experienced directly through a set of complex social situations. He also draws on the example of infants, who, by the time they start using these and similar notions in language, have already acquired "implicit understandings" of them (Barsalou, 1999, p. 602). An interesting extension of the idea of how perceptual symbol systems can be used when dealing with abstract concepts and metaphors can be found in Gärdenfors (2004, 2014). As Gärdenfors puts it, although physical and abstract domains are separated from each other, they do not exist in isolation. This means that the structure from one domain can be "imported" into a different one as long as two concepts have at least one common domain or dimension. However, this is not always the case. For instance, in the case of metaphorical extensions, the mappings can still be created between the non-related domains. An example that Gärdenfors gives concerns the metaphoric expression "bumpy relationship". Analogous to the notion of a "bumpy road", which is represented in a form of two axes, where x is length and y is height, "bumpy relationship" is projected into two different spatial dimensions: time and problem level. The resulting graph is similar in its form in both cases, which allows for the metaphorical extension to happen (see Figure 2.2)

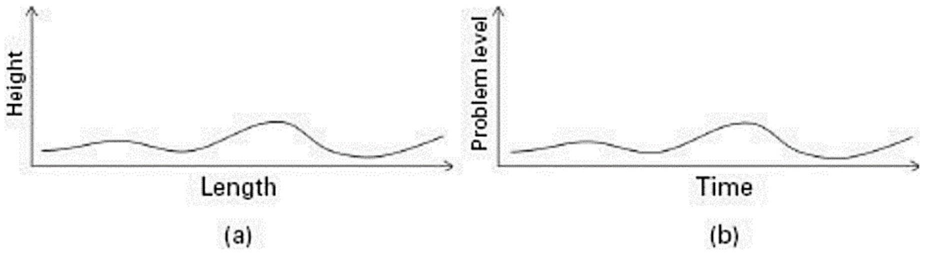


Figure 2.2. Spatial projection of physical domain to abstract (Gärdenfors, 2014, p. 224)

Such analogies based on the geometry of meaning provide a principled explanation of the emergence of conceptual metaphors in gesture and language, where the source domain mapped onto the target domain is usually grounded via sensorimotor experiences, as will be shown in Section 4.3.1. (Metaphor). Moreover, the idea behind productivity highlights the importance of spatial relations for cognition and language. It is yet another important topic in gesture research, as gestures (as well as manual signs) referring to both concrete and abstract notions are produced in a three-dimensional space, thus using elements of perceptually built knowledge. The importance of special relations has been pointed out previously in sign language research, which showed that experience with sign languages enhances people's visual-spatial skills (e.g. better performance when carrying out mental rotation tasks (Kubicek & Quandt, 2021)).

## 2.2.3 Modality dependence

Probably one of the most important features of PSS for explaining the nature of gestures is the correspondence of their structure to the referent. In other words, due to selective attention, "symbols can operate on any aspect of the perceived experience" (Barsalou, 1999, p. 585). This has two crucial implications. First, the modality-dependent view on perceptual symbols (unlike the amodal view) allows for assuming the existence of modality-conditioned symbols in the long-term memory, which can be retrieved and activated depending on the communicative need at hand. Secondly, such approach also accounts for variability in perception. This is especially interesting, though, when talking about the nature of iconic signs. Although all humans have a shared general way of processing perceptual information, the individual nuances in the perception of shapes, textures, space, etc., together with the attentional shifts, lead to variability in expressing this information.

It is worth mentioning here that the visual system, albeit powerful, is not the only source of information about the world. This work pays special attention to visual perception due to its direct connection to bodily modality and gesture (gestures are experienced through vision); however, our knowledge of the world is not restricted to what we perceive visually. For example, such areas of research as instrumental actions and their role in the evolution of language, hierarchical organization, and the use of gesture have come to the fore in recent years (e.g. see Novack & Goldin-Meadow, 2017). Although this and similar theories will not be discussed in detail in the current work, it is important to emphasize that both the PSS model outlined in this chapter, as well as the framework of Multimodal Parallel Architecture, which this thesis aims to elaborate are easily adaptable and flexible enough to be extended to accommodate other models of perception and conceptualization.

## 2.3 Conclusion: implication for co-speech gesture research

The multimodal view of the conceptual system bears importance for both language and gesture. All the theories discussed in this section can prove useful to account for both gesture production and recognition as they make the first step in explaining how humans might conceptualize the world in a modality-dependent way and how gestures can represent the conceptualized objects. One of the theories, Barsalou’s theory of PSS, is especially useful for explaining how human’s sensory and perceptional systems guide the conceptualization process, which, in turn, allows for stored mental representations to be activated in different tasks, one of which is communication and using language. Barsalou’s theory attempts to explain how any type of information about objects and events acquired through the mechanisms of perception, proprioception, and introspection gets stored in the long-term memory to be retrieved when needed.

In order to capture any visuospatial relations, the human brain has to employ coordinate systems. These coordinate systems are constructed to account for the specific viewpoint to provide viewer-centered or object-centered representations. The representations, in turn, are built based on sets of primitives, “the most elementary units of shape information available” (Marr, 2010, p. 300), such as location, size, orientation, etc., associated with stable geometrical characteristics. Moreover, the schematic character of representations stored in a long-term memory enables the grouping and categorization of perceived concepts based on certain features. This leads to the componential nature of symbols, which, together with selective attention, allows to bring certain recognizable features of objects to the foreground, thus making some recurrent aspects of the representation more salient. In line with the theory of conceptual spaces, the primitives can be seen as analogues of primitive dimensions, some of which can be naturally expressed in gesture. They also correlate with the parameters of gesture (gesture space, orientation of the palm, etc.) outlined in (section 3.2 Kinesthemes) and can explain the emergence of stable form-meaning associations characteristic for recurrent gestures (see section 3.1 Recurrent gestures and gesture families) and potentially, for all kinds of spontaneous referential gestures. It is possible to claim that gestures can naturally capture a number of domains such as size, shape, width, volume, weight (for nouns); force, direction, path (for verbs); (relative) location (for prepositions, demonstrative pronouns); manner (for adverbs); etc. Hence, gestures have the potential to emphasize certain domains that are not treated as salient in speech. For example, the concept of “apple” has at least four domains: color, taste, shape, size, and nutrition (after Gärdenfors, 2022). While all these domains are indeed grounded in perception, the domains of color, taste, and nutrition can be only naturally expressed in language, the domains of shape and size can be represented in gesture even when absent in speech. The bodily modality can be argued to pertain to those domains, which can also explain the metonymic nature of bodily movements: only one or two domains or dimensions can simultaneously stand out to represent the concept as a whole.

It is also worth mentioning that the use of gestures is dynamic and is not restricted to encoding static information. In this, the theory based on the proposals of Barsalou (Barsalou, 1999) and Hafri (Hafri et al., 2023) has an advantage over other theories. As will be seen in the next chapters, co-speech gestures represent objects not only by means of modelling but probably more frequently by interacting or tracing. Additionally, gestures can capture movement patterns and manner of motion or, more broadly, function as attributes in multimodal constructions (for details, see Chapter 5 Syntax).

Lastly, humans' sensitivity to spatial relations and their ability to subject mental representations to transformations play an important role in the use of gestures and manual signs, as the research on sign language shows (for details, see e.g. Kubicek & Quandt, 2021; Secora & Emmorey, 2020)

In sum, the argument is that PSS is an autonomous system mainly derived from our sensory experiences with the world, which can interface with semantic and syntactic systems in language. Moreover, the perceptual basis must be a reason for the presence of both impressive similarities and crucial differences between bodily and auditory modalities, which are yet to be described and researched in full. The implementation of these models is limited in the case of fully developed complex sign systems like spoken or sign languages due to the higher degree of conventionality and arbitrariness of their signs; however, they give insight into the evolution of language and lay the theoretical ground for the theories regarding the nature of gestures that emerge later such as the theory of image schemas, metaphor, and metonymy. These theories will be discussed in the following chapters.

# 3 Morphology

Initially, co-speech gestures were considered to be non-conventional iconic or indexical signs. In other words, they used to be understood as tokens (but not tokens of types) or so-called 'singularities' (Kockelman, 2005), that can become signs only in the presence of verbal context. Hence, as Enfield (2013, p. 697) puts it: "the problem of comprehending gesture meaning is taken to be one of interpretation (from token form to token informative intention)". However, contrary to popular belief that co-speech gestures are idiosyncratic elements that cannot be lexicalized or grammaticalized to establish stable form-meaning relations across various contexts required to conform to the accepted view on morphemes in linguistics, current research on gesture families and gesture morphology has shown that the structural organization in gestures is complex and resembles that of fully-fledged spoken and sign languages. Importantly, by drawing parallels in this and the following chapters between the structure of co-speech gestures and fully developed languages, we focus on the fact that *gestures are analyzable and structurally complex and decomposable entities which do not (only) feed off more elaborate language systems*, but also possess features of their own which are *modality specific*.

In order to build an argument, we should adopt an accessible terminological apparatus. It's essential to recognize the wealth of knowledge gathered through years of studying language and gestures. Ignoring this foundational knowledge would be unwise. Thus, our central aim is to elucidate the existing research on the interplay between gestures and spoken language before discussing the strengths and weaknesses of the presented views.

In this chapter, the concepts of recurrent gestures and gesture families are introduced following by the suggestions of what theoretical and experimental apparatus for the analysis of gestural forms can be deployed in order to establish evidence that single instances of gesture use (both the single tokens and complex multi-type gestures) have an inherent complexity and compositionality of form based on the four parameters: gesture salience/space, the orientation of the palm, hand shape, and movement. This would provide insights into the system of morphology in gesture *comparable* to that of spoken language, where morphological organization follows certain generative principles that allow for interfacing with other language systems.

## 3.1 Recurrent gestures and gesture families

The first attempt to structure and categorize co-speech gestures based on shared semantic domains developed into the theories of recurrent gestures and gesture families. These concepts marked the shift from seeing gestures as isolated entities to looking for the interrelations between gestures and their form/meaning parameters. Initially proposed by Adam Kendon (2004), the theory of gesture families and recurrent gestures found its application in the works of (Fricke, 2014a), Bressem & Müller (2014; Müller, 2004), Ladewig (2011, 2014, 2020), Teßendorf (2014), among others.

Recent research has shown that silent gestures undergo processes of schematization and stabilization and can become a part hierarchical system and have the potential to form structural wholes. Effectively, recurrent gestures and gesture families represent bottom-up and top-down approaches, respectively. In other words, the researchers look at the co-speech gesture from two perspectives: semasiological, to analyze gesture families based on the similarity of form, and onomasiological, to adopt a broader view of gestural fields centered on shared semantics without necessary resemblance in form. Recurrent gestures are the ones singled out by the similarity between their forms. They were the first to fall into the focus of researchers due to their salience and pervasiveness. Although initially different terms to define such gestures were coined by different investigators (e.g. "catchments" (McNeill, 2005), "pragmatic gestures" (Kendon, 2004), "interactive gestures" (J. B. Bavelas et al., 1992), etc.), all of them pointed out two important features of this type: a certain degree of conventionalization and cultural affiliation. Recurrent gestures were the first step that allowed for arranging a seemingly unordered variety of bodily behaviors into a number of organized sets of gestures with rather stable form-meaning pairings ("G-family" and "R-family" (Kendon, 2004); "away gestures" (Bressem & Müller, 2014b); "cyclic gesture" (Ladewig, 2014b), etc.). Despite their conventional character, these gestures cannot be considered emblems for two reasons. Firstly, although restricted to a limited number of uses, their meaning largely depends on the context and their structural position in the utterance. Secondly, the meaning of recurrent gestures can be inferred from their form: albeit schematic, they still bear resemblance with the actions and instances they have been derived from. As a result, researchers working on co-speech gestures started arranging databases of gestural repertoires typical for particular cultures (e.g. repertoire of German recurrent gestures (Bressem & Müller, 2014a)). Subsequently, it was observed that gestures within two sets of recurrent gestures can display similarities in meaning without necessarily sharing the same form parameters. Importantly, the number of gestures that share the same functional characteristics is also limited and allows for a different approach to categorization, which resulted in the emergence of the related concept of gesture families.

Nevertheless, there was a major challenge related to these types of analyses. When analyzing gestural forms, researchers were guided by their own judgements regarding what differentiates gestures within one set from gestures which do not belong to this set. In other words, while referring to a set of recurrent gestures as sharing resemblance in form, the focus, in fact, remains on the most salient feature(s) of this form. That is, in the case of recurrent gestures, it is not required that all their instances across all contexts are identical; rather, they convey local meanings by employing additional parameters either idiosyncratically or in a more systematic manner (for an example of Cyclic gesture see Ladewig, 2014b). These additional parameters can be established by analyzing, based on an adaptation of four parameters used in sign language research (Bressem, 2013), gesture salience/space, the orientation of the palm, hand shape, movement, and measuring the level of dissimilarity between gestures, i.e. kinematic entropy, based on these features. For example, when the entropy is lowest, it means that interrelationships are distributed in a more systematic way. Subsequently, if a high level of systematicity and typification is established, it makes it possible to define variants of a recurrent gesture. This provides an answer to a number of questions related to the contextual differences between recurrent gestures and the nature of gesture families. In other words, particular parameters of gesture form (not gestures as a whole) establish stable form-meaning connections and then become the core of several gestures forming a gesture family. These findings allow some researchers to argue that co-speech gestures possess rudimentary morphology (Fricke 2013, Müller 2004) or "emerging morphosemantics" (Kendon, 2004, p. 224). Moreover, as Ladewig notes: "increase of stabilized clusters of form parameters goes hand in hand with the process of lexicalization, grammaticalization or pragmaticalization" (Ladewig, 2020, p. 30). Such "variations in form and function point to a rudimentary gesture morphology that structure this small-scale gesture family" (Müller, 2004, p. 254 as cited in Fricke et al., 2014, p. 1633). For example, the research on a recurrent Palm Up Open Hand gesture (PUOH) has shown that "based on formal and functional variations of the kinesic core through various movement patterns (rotation, lateral movement, up and down movement), the semantic core of offering, giving, and receiving objects is extended to mean "continuation", "listing ideas", and "a sequential order of offered arguments or presenting a wide range of discursive objects" (Müller, 2004, p. 254 as cited in Fricke et al., 2014, p. 1634). Another example illustrates similar properties of a Cyclic gesture. The core semantic feature of this recurrent gesture is the representation of the concept of continuity. The formational core of this gesture, in turn, consists of the rotation of the wrist or any other segment of the arm. However, it is used to convey a number of additional meanings, such as signaling word or concept search, describing, or requesting (Ladewig, 2014b). All these additional layers of meaning are reflected in the changes of movement pattern and/or form parameters, such as the use of particular section in a speaker's gesture space or the utilizing larger movements as shown in Figure 3.1.

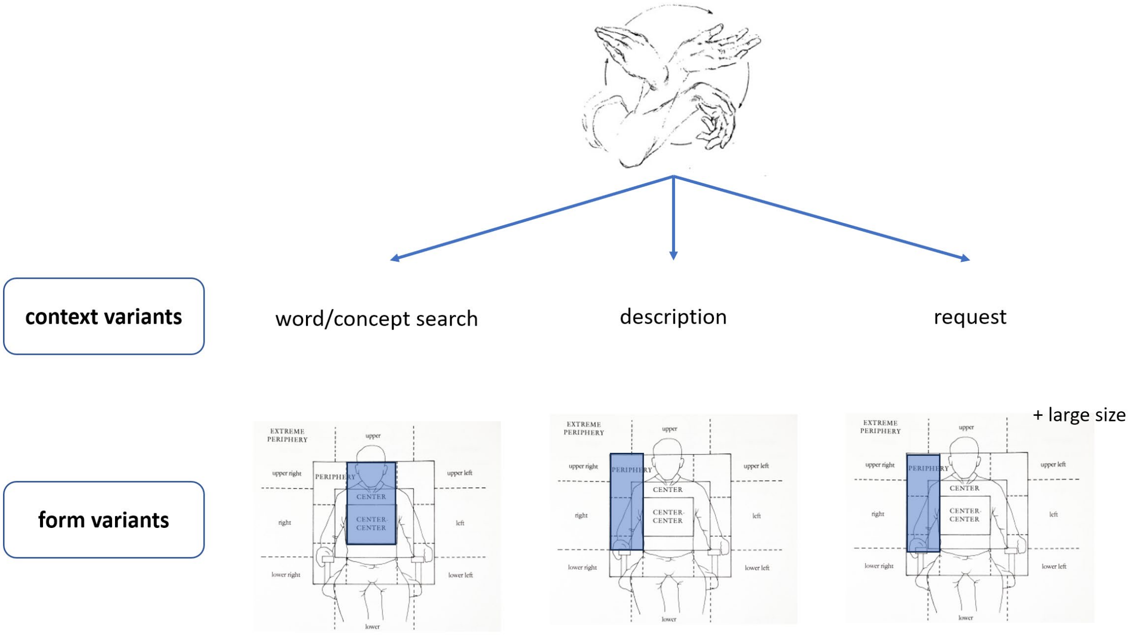


Figure 3.1. The use of Cyclic recurrent gesture (after Cienki, 2023)

Interestingly, gesture families can also form hierarchical relations when gestures comprise one family which already constitutes another gesture family. For instance, the family of "away gestures" together with POUH gestures are included in the family "gestures of negation". Thus, it is possible to capture complex interrelations that exist in co-speech gesture similar to those between words in a spoken language.

There is still, however, the lack of empirical evidence for the absence of commonly accepted procedure for the analysis of gestural forms. The two promising solutions are described in the following sections: kinematic analysis proposed by Wim Pouw and his colleagues (Pouw, de Wit, et al., 2021; Pouw, Dingemanse, et al., 2021) (Section 3.3) and kinesiological approach developed by Boutet et al. (2021) (Section 4.1). Both these approaches explore the biomechanics of human body using different types of motion capture techniques to develop an objective view on how the gesture is originated in the body and, thus, refine the existing methodology used for the analysis of the parameters of gestural forms. These technologies are likely to provide unambiguous information on the core elements of recurrent gestures and gesture families and make a step away from trying to accommodate co-speech gestures into the existing theoretical frameworks (primarily focused on analyzing the structure of spoken language) and treating other modalities (in particular, manual-visual) as subsidiary and entirely dependent on an elaborate system of spoken language.

## 3.2 Kinesthemes

A notion congruent with the idea of gesture fields and recurrent gestures is the notion of kinestheme elaborated by Elen Fricke (2014). Fricke argues that in order to explore potential typification and systematicity in co-speech gesture it is useful to look at these processes beneath the morphological level but above the level of single sounds and their combinations. Based on her observations, Fricke contends that co-speech gestures have stable underlying semantic relations. To define these relations, she uses the term 'kinestheme' by analogy with the term 'phonestheme' which reflects the "processes of morphological contamination or blending in word formation of spoken languages" (Fricke, 2013, p. 741). Fricke claims that the notion of kinestheme in gesture "supports and further elaborates the hypothesis of a "rudimentary morphology" in co-speech gestures" (Fricke, 2014a, p. 1619).

According to Fricke, phonestheme in speech and, as a result, kinestheme in gesture, can be defined as "a set of semanticized submorphemic tokens whose similarity on the level of form correlates with a similarity on the level of meaning" (Fricke, 2014, p. 1620). She explains this concept by using Zelinsky-Wibbelt analysis (Zelinsky-Wibbelt, 1983) according to which submorphemic structures are analyzed according to the semantic features of the words in which these structures occur. Then, the set of structures is formed based on the shared features. In spoken language, this results in defining phonesthemes such as *sm-* that occurs in words like *smoke*, *smog*, *smirch*, etc., which is associated with the concept of dirtiness. The same trend was also observed in gesture: for example, the set hand shapes used by German speakers as non-iconic pointing gestures can differ in form while, at the same time, sharing certain parameters (e.g., extension of at least one finger in the direction of the referent, occurrence in gestural space away from the speaker. etc.) as illustrated in Figure 3.2.

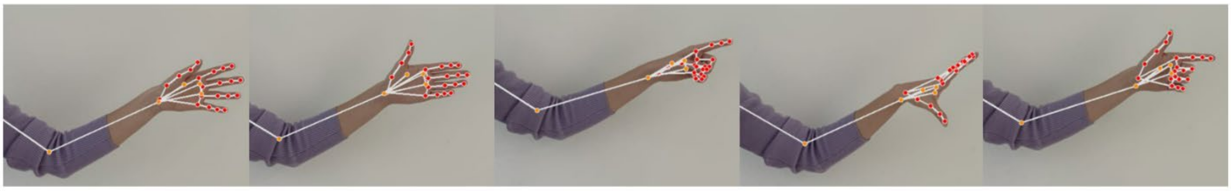


Figure 3.2. Family of PLOH (palm lateral open hand) + G-form as non-iconic pointing gestures.

Importantly, the theory draws on Peirce's notion of diagrammatic iconicity which can account for the direct similarity relationships between the signs. As Willems and De Cuypere (2008) point out "diagrammatic iconicity in language is more than form-meaning isomorphism; it is a structure inherent in the verbal form itself irrespective of whether the diagram is used to represent anything at all" (Willems & De Cuypere, 2008, p. 73). In other words, diagrammatic icons are mental diagrams irrespective of whether they are graphically represented or not. Hence, form-meaning relations of this type are "relatively motivated", and, as Saussure argues "the more a sign system is relatively motivated, the more grammatical it is'' ( Saussure, 2005, p. 133, as cited in Fricke, 2014, p. 1621) Fricke also uses the concepts *semantic loading*, which means the semantization of submorphemic elements based on common semantic features (family resemblances).

In her analysis of gestural kinesthemes, Fricke deliberately avoids using iconic gestures to prevent explaining semantization processes as based solely on iconic relations between the gesture and the referent. Instead, she focuses on non-iconic deictic gestures, which, although represented by a number of different handshapes, share the same context-independent functions.  Moreover, Fricke distinguishes between the two types of kinesthemes: simple and complex. While simple kinesthemes are defined as "intersubjectively typified and semanticized gestural tokens whose similarity on the level of form correlates with a similarity on the level of meaning" (Fricke, 2014a, p. 1618), complex kinesthemes represent the next step in the process of word formation, namely "morphological contaminations'' (Fricke, 2014a, p. 1624). For instance, Ladewig (2011) suggests that the core parameter of the PUOH gesture can be fused with the core parameter of the Cyclic gesture resulting in a gesture that describes the continuous listing of ideas. Thus, such compounds unlock the potential of kinesthemes to possess rudimentary compositionality in gestural morphology and semantics.

The concept of kinesthemes can be extended to account for the existence of variety in recurrent gestures described in Section 3.1 (Recurrent gestures and gesture families). To put it another way, kinesthemes can be used as a term to represent the core and additional parameters in recurrent gestures. *This approach goes in line with the theory of gesture families; however, it provides a more principled and coherent way to analyze and systematize various instances of recurrent gestures used by speakers by also reflecting their unique status and pointing out structural differences between co-speech gestures and speech*.

## 3.3 Kinematics of gesture

As has been seen in Sections 3.1 (Recurrent gestures and gesture families) and 3.2 (Kinesthemes), most of the research aimed at uncovering morphological semantics in gesture used observation as a research method. However, the most recent research by Wim Pouw and his colleagues from Max Planck Institute for Psycholinguistics carried out a series of experiments in order to test whether the relations between categorical linguistic content and continuous kinematics can be established (Pouw, de Wit, et al., 2021; Pouw, Dingemanse, et al., 2021). Instead of studying gestures by inferring meaning from their observable form, which dominates the field of gesture research, for the time being, the researchers focused on the kinematics of each gestural form in relation to other forms. This approach is derived from research on lexical semantics, which uses among others the framework of componential analysis to establish systematicity of relations between lexical items based on their semantic features. Unlike previous studies that focused largely on the top-down approach, which starts with defining larger domains and then narrowing them down to functions, in their attempts to establish interrelations between gestures, the experiment conducted by Pouw et al. used the bottom-up approach, whose primary focus is to determine the status of individual linguistic elements. i.e. assign functions to linguistic elements before moving on to the domain level.

In keeping with the previously described theoretical stance, the researchers also state that gestures are not mere singularities[5] but have an underlying fixed form-meaning relationship / form-meaning mapping on the kinematic level. To some extent, gestures have similar to language systematic properties with various degrees of variance. Gestures can also be used as a bootstrapping mechanism due to their permanent traits. Analogous to semantic similarity space determined for lexical items, gestures form kinematic similarity space. Kinematic space co-varies with semantic space, semantic distances between lexical items can reliably predict kinematic distances in gesture use, thus, there is a positive relation between semantic and kinematic distances of both within- and between-category items. A bottom-up approach to building semantic maps has shown that the semantic proximity of concepts within the same semantic cluster is in direct relation to the lower kinematic distances, while their comparison to concepts outside a given semantic field causes an increase in kinematic distances between gestures. It is important to mention, however, that it does not mean that gestures and lexemes that represent concepts from the same semantic field scale with the kinematic distances between those concepts.

One of the studies carried out by the research group presented the analysis of the 1200 recordings of silent gestures, representing 24 concepts with various degrees of semantic relatedness and proximity produced by 98 participants in 6 "generations" (Pouw, Dingemanse, et al., 2021). The researchers focused on the set of five kinematics properties to discover connections between the gestural forms produced within and between the "generations" of participants. They discovered that gestures become "on average smaller, less temporally variable, and less intermittent as the communicative system matured" (Pouw, Dingemanse, et al., 2021, p. 16). The results showed that the "younger" the "generation", the more reduced entropy, "the amount of information that is needed to compress the signal set" (Pouw, Dingemanse, et al., 2021, p. 5), the kinematic networks have, which indicates the increased systematicity in gesture use. The study also revealed that although the gesture system overall displayed more coherence, gestures depicting semantically related concepts follow their own developmental paths over generations diverging from other lexical chains. These two processes denied the possibility that compression in gestures is solely driven by the tendency towards communicative efficiency or minimization of communicative effort, as "communicative efficiency does not automatically entail systematicity" (Pouw, Dingemanse, et al., 2021, p. 22). Importantly, the overall development of silent gestures displayed a tendency similar to that of the fledgling sign languages (such as ISN[6]) and spoken languages in general, where the processes of compression and systematization lead to the gesture segmentation with the gradual emergence of conventionalized functional markers aiding the process of disambiguation, which, in turn, increases the potential for combinatoriality or generativity of the elements of the system. To sum up, silent gestures transferred over several "generations" of participants demonstrated the potential of silent gestures to comprise a linguistic system that manifests itself in communicative tokens and gradient and

[5] tokens but not tokens of types

[6] Nicaraguan Sign Language/Idioma de señas de Nicaragua (for more details see, e.g., Senghas & Coppola, 2001)

continuous relationships between them. The study of the aspect of form in silent gesture carried out by Pouw W. et al. showed the gradual shift from iconicity in gesture to comparative idiosyncrasy and systematicity across gestures which tells about the presence of linguistic constraints placed on the kinematic system.

It is important to emphasize that the experiment did not analyze spontaneous co-speech movements, but pantomime (silent) gestures. This means that the results might be not consistent with the findings if spontaneous hand gestures are used. It is possible, however, that replicating the experiment using spontaneous bodily movements can provide experimental explanation for the theory of gesture families and the notion of recurrent gestures and kinesthemes. If such experiments show that co-speech gestures comprise stabilized clusters of form parameters, it will be possible to argue that co-speech gestures are not just spontaneous movements, but indeed undergo processes of lexicalization, grammaticalization, or pragmaticalization. These findings are also in line with the widely accepted view on the course of language evolution where the processes of historicity and adaptivity play the central role, since the increasing communicative efficiency augments the learnability and comprehensibility of the gestures, while communication is an act of continuous and dynamic process of sender-receiver co-adaptation.

## 3.4 Conclusion: insights and future directions

As has been shown in the chapter, co-speech gestures possess similarities with morphological structure of language on the level of form. This can help to explain which principles generate the system speech-entangled gestures and how this system can be recruited for communicative purposes. However, *not all the features of the system of gesture can be comparable with those of a spoken language as has been shown in the research on kinesthemes*. Thus, in order to further explore the morphological potential of co-speech gestures and their ability to undergo the processes of lexicalization, grammaticalization and pragmaticalization, new methods of analysis and new parameters have to be developed. In this chapter, one of such methods, the kinematic-based analysis of gesture (Pouw, de Wit, et al., 2021; Pouw, Dingemanse, et al., 2021) was discussed. Nevertheless, some other possible approaches will also be mentioned in section 4.1 (Gesture classifications) of the next chapter.

# 4 Semantics

The issue regarding whether spontaneous bodily movements and, in particular, co-speech gestures have meaning of their own or they should be seen as instances or tokens that can be analyzed only together with speech has been debated since the first studies of multimodality appeared in the second half of the twentieth century (Birdwhistell, 1971; Kendon et al., 1975; Pike, 1967). There is still no clarity in where co-speech gestures stand in term of language-based perspective on meaning; however, their integral role in speech is rarely disputed, and more evidence appears that gestures have an internal semantic structure and conceptual hierarchy of their own. It is important to note that reference does not entail direct mappings between the verbal or gestural expressions and the physical world; rather, the expressive power of various means depends on humans' conceptualization of the world. The current approach to the meaning-form relation of co-speech gestures uses cognitive-semiotic principles of iconicity, indexicality, metaphor, metonymy, and image schemas in order to explain the motivation behind the structuring and functioning of multimodal messages and the way semantics and pragmatics operate across modalities. Although spontaneous co-speech gestures do not show levels of symbolicity and conventionality typical for symbolic sign systems (spoken language, sign language, etc.), they display consistent behavior with respect to similarity, contiguity, and schematization. The framework stems from the idea that the internally motivated semiotic structure is "inherent to communicative gestures" and bodily movements in general (Mittelberg, 2013). In other words, all gesticulations (in our case hand movements co-occurring with speech) are motivated/intentional and meaningful and can be characterized by their "deliberate expressiveness" (Kendon, 2004, p. 15).

Importantly, meaning is argued to be not static, arbitrary, and abstract, but rather dynamic, motivated, and concrete; it is argued to arise through a process (McNeill, 2005), which highlights the importance of enchronic analysis, i.e. the analysis of conversational elements in their relation to adjacent instances in a coherent sequence of interaction as opposed to diachronic analysis. A slightly modified approach to meaning was proposed by Mittelberg, who claims that the meaning arises "in the dynamic gestalt of a mental representation or some other kind of cognitive and/or physical response to a perceived sound, word, image, or human behavior" (Mittelberg, 2013, p. 759). She also emphasizes that "the body portrays, i.e. exbodies, how the person conceptualizes and understands the abstracta" (Mittelberg, 2013, p. 776).

Recent research on sign languages has highlighted the importance of humans' physical experiences with the world in the emergence of signs. Virginia Volterra and her colleagues (Volterra et al., 2022) who carry out research on Italian Sign Language (LIS) as a part of an Institute of Cognitive Sciences and Technologies of the National Research Council group (ISTC of the CNR) draws parallels between co-speech gesture and signs in sign language. In the book *Italian Sign Language from a Cognitive and Socio-semiotic Perspective*, researchers emphasize the shift of focus in sign language studies toward exploring the importance of gestures in the process of language acquisition in both hearing children and those learning sign languages. The CNR group found out that a lot of previous research on sign languages often dismissed gestures as the first stage in language development interpreting them as fledgling signs rather than gestures; however, it turned out that both hearing and deaf children follow similar developmental trajectories before they produce their first words/signs. Firstly, gestures enter every child's repertoire at pre-linguistic stage. Among those gestures, researchers typically distinguish deictic and other performative gestures, gestures representing intransitive actions, and gestures that iconically depict manipulation of objects (i.e. representational gestures). Thus, it is possible to see that the types of gestures children start using early on already reflect the typology established for the purpose of describing and classifying co-speech gestures. Subsequently, at later stages of language acquisition, children still tend to resort to gesture to accompany or substitute words, which leads to the production of so-called crossmodal combinations. Interestingly, both hearing and deaf children abide by similar motor constraints as they use the same handshapes when producing gestures. This shows "a clear continuity between gestures used in spoken language, those used without speaking, and signs" (Volterra et al., 2022, p. 19). Thus, the existing body of research on both gesture and sign language proves "the central role of the motor system in the construction of meaning" (Volterra et al. 2022: 18). Moreover, it makes it possible to establish connections between early gesture use and the speech-gesture system evolved in adults. It is also important to note that Volterra et al. argue that both co-speech gestures and signs are grounded in actions, thus corroborating the claim that human's expressive power is tightly interrelated with our embodied experiences with the world. As Volterra et al. put it: "gestures accompany the child in the transition from communication strongly anchored within a specific context to the complete decontextualization of the first linguistic forms, fulfilling a complex role of interface between action and the early development of language" (Volterra et al., 2022, p. 18).

Unlike Chapter 2 (Spatial Structure and Conceptual Structure) which focuses more generally on the relation between perception and cognition and only its potential extension to the theory of communication, in this chapter the focus is on a number of cognitive linguistic phenomena, which originally were treated as properties of verbal language and its relation to perception, and only later were adopted by researchers on sign language and co-speech gesture. That is, the main difference is that in this chapter, the notion of schemas is looked at as a phenomenon linking conceptual system with the system of language, while when talking about CS and SpS the discussion focused on the connection between perceptual mechanisms and mental representations. The chapter starts with the section describing and analyzing various approaches to the classification of gestures in order to create a terminological apparatus that will be useful for the subsequent discussion of both semantic and syntactic phenomena. Next, four sections on image and mimetic schemas as well as metaphor and metonymy will follow where the similarities and differences between spoken language and co-speech gestures will be stated explicitly.

## 4.1 Gesture classifications

In comparison to verbal forms, where form-meaning relations are seen as arbitrary and based on convention established among the speakers of a given language, kinesic forms in gesture are less independent and can be frequently determined by aspects of their meaning. Moreover, inferring meaning from gestural form depends on conversational settings or 'gesture ecologies' (Streeck, 2009, 2013), i.e. relations of gestures to the physical context in which they occur. Even when gestures undergo a process of conventionalization "their form-meaning relation is motivated and the motivation is still transparent” (Müller, Ladewig, et al., 2013). Thus, the interpretation of gesture meaning relies on understanding of their form and/or their function irrespective of their status as conventional signs. Importantly, these two parameters are often treated as complementary, which creates an inherent problem of form and function conflation within proposed sets of categories, as pointed out by Müller (1998) and Cienki (2022). This multidimensionality of gesture form-meaning relations brings together a number of approaches to meaning, such as semiotic, pragmatic, cognitive, and interactive.

Peirce’s triadic dynamic theory (Peirce, 1955) is often considered a starting point in the analysis of multimodal data. Standard gestural taxonomies, such as the ones proposed by McNeill (1992) and Kendon (2004), as well as more recently developed gesture typologies are largely explicable from the semiotic perspective. In line with the Peircean theory, gestures, similar to any other signs, are grouped into three categories based on their semiotic function: indexes, icons, and symbols/emblems (Fricke et al., 2014; Mittelberg, 2008). In this interpretation co-speech gestures are functionally compared to verbal signs in their ability to represent something other than themselves. Symbols/emblems are typically excluded from the analysis of spontaneous bodily movements because they are characterized by high degree of conventionalization, arbitrariness, and cultural loading, which sets them apart from the rest of the co-speech gestures. Indexes were initially presented in a form of deictic gestures and narrowed down to pointing gestures (McNeill, 1992); however, as will be seen, the notion of indexicality in co-speech gesture spans across a number of gesture types and proves to be much more abstract. Although indexes are grounded in time and space, their gestural (and verbal) forms do not refer to the entities which are physically present in the communication setting: indexed referents can be present only in the conceptual space shared by the speakers (abstract deixis). In other words, deictic gestures point at the ideas mapping them onto the real space. Iconic gestures (sometimes the term ‘referential gestures’ is used) “establish reference to either some object, person, property, or event, especially for talking about some kind of physical topic” (Cienki, 2017, p. 93) and are characterized by high variability of forms and constitute probably the most versatile group that deals with an array of representational and pragmatic functions. This led to the emergence of more detailed gesture taxonomies, where structural resemblance between form and meaning is signaled, for instance, through different acts, such as tracing, molding/holding, enacting, embodying, etc., which highlight the certain features of the referred concept (representational function). Another typology proposed by Mittelberg (2014) distinguishes three subtypes of iconicity: image, diagram, and metaphor. It captures an important fact that iconic gestures do not only refer to the physical entities, but also to abstract notions (entities, properties, processes) by means of metaphor. This typology will be discussed in more detail in Section 4.3 (Metonymy and metaphor), as both metonymic and metaphorical extensions can be seen as internal cognitive strategies inherent in the process of sign/gesture formation. Alternatively, the so-called ‘pragmatic gestures’ (Kendon, 1995, 2004) which serve as markers of speech acts or aspects of discursive structure. As Cienki (2017) explains it, discourse-related gestures are related to the discourse itself or refer to the topic and/or speaker’s attitudes towards it (e.g. using palm up open hand (PUOH) gesture with both hands to express the idea “I’m not sure/I don’t know” as shown in figure 4.1). The first type of discursive gestures is used to structure the discourse or the argument by referring to the ideas as placed in space or they refer to the topic being discussed; however, one of the examples that Cienki (2017, p. 93) gives concerns “point[ing] to different spaces coordinating with mention of ideas”, which means that these gestures can be an extension of a broad category of pointing gestures with abstract deixis. Such multiplicity of interpretation does not allow to solve the problem of gesture categorization completely and definitively outline the boarders between the gesture types. Hence, within one typology, different strategies for identifying gesture types are employed which makes gesture types not mutually exclusive. For example, interacting gestures (Enfield, 2013) can be both iconic, as they imitate actions, and indexical, as the hand shape is not the shape of a referent, but determined by it. However, not all the taxonomies are characterized by their increasing complexity: for instance, some of the more recent classifications suggested by Müller (2014) and Cienki (2013) significantly reduced the scope of gesture types and subtypes. Overall, although being guided by similar theoretical bases, there is no unanimity among researchers on how to consistently define gesture types.

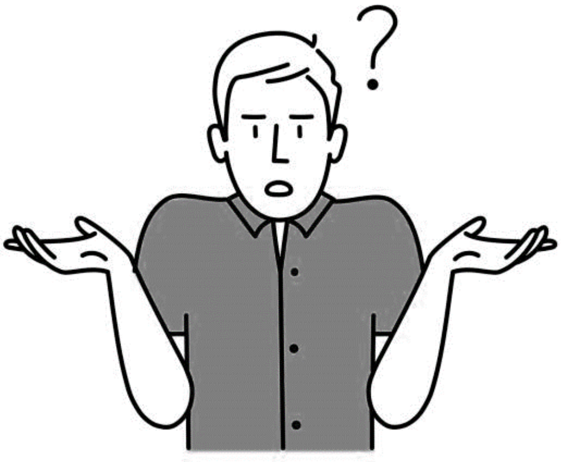


Figure 4.1. PUOH (palm up open hand) gesture used to with both hands to express the idea "I'm not sure/I don't know"

Recently, two modern approaches to the analysis of gesture emerged in the field of gestural studies. Both of them are still in the process of development and, thus, will be discussed briefly solely with the purpose of showing what alternative methods can be used in the analysis of co-speech gesture and what immediate practical application they have. One, called multi-vector semiotic model (Iriskhanova & Cienki, 2018), focuses on the variety of semiotic functions performed by gestures; the other, termed kinesiological approach (Boutet et al., 2021), is built around the biomechanical properties of human body.

Although researchers aim to create clear and simpler accounts on gesture typologies, most of the existing classifications fail to account for multifunctionality of gestures and the absence of clear-cut distinctions between different types of gestural signs and their functions. Iriskhanova and Cienki (2018) revisited the notion of continuum proposed by Kendon (1988) and revisited by McNeill (2005) who proposed that instead of assigning gestures to discrete categories, the form-meaning relations in gestures are better represented as a continuum based on the status of gestures as conventional signs with arbitrary idiosyncratic gestures at one end and highly conventional emblems at the other as shown in Figure 4.2.

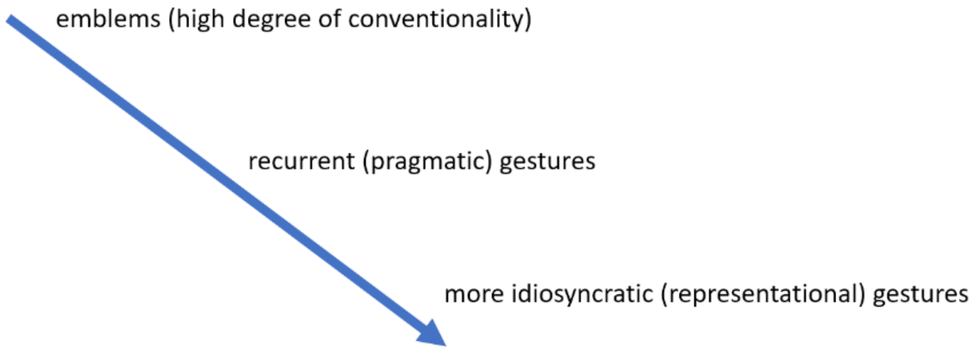

emblems (high degree of conventionality)
recurrent (pragmatic) gestures
more idiosyncratic (representational) gestures

Figure 4.2. Adapted version of Kendon’s continuum

Iriskhanova and Cienki argue that the idea of continuum now also extensively used in cognitive linguistics when talking about semantics-pragmatics and grammar-lexicon continua is the best way to reflect the complicated nature of gestural forms and functions which go beyond classically proposed typologies outlined above. As a solution to the problem of gesture classification they propose a multi-vector model instead of mono-dimensional continua based on the array of semiotic features that are mentioned in various studies of gestures: "conventionality, semanticity, arbitrariness, pragmatic transparency, autonomy, social and cultural import (symbolism), awareness, recurrence, iconicity, metaphoricity, indexicality, salience" (Iriskhanova & Cienki, 2018, p. 30). The researchers propose using a diagram with a grid pattern surrounding twelve vectors reflecting the abovementioned features as shown in Figure 4.3. In this Figure, the way of analyzing emblems is presented by assigning points to each of the parameters from 0 to 3 (low to high) to reflect the approximate degree of presence of each semiotic feature in a given gesture.

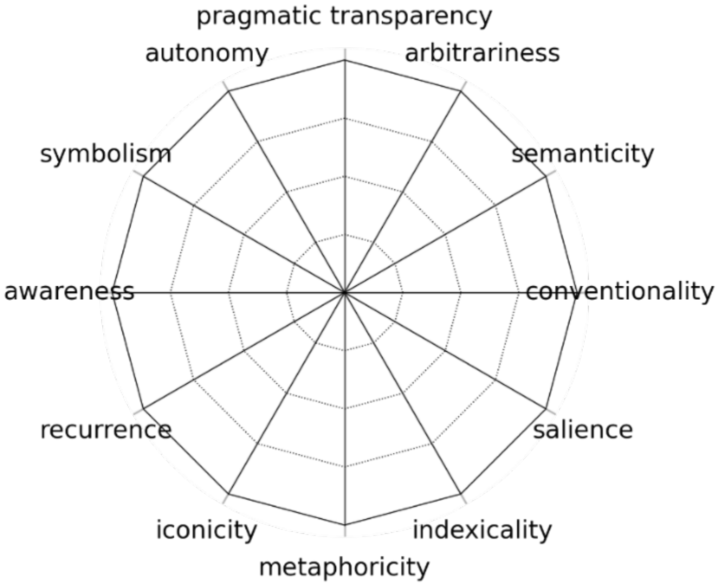


Figure 4.3. Multi-vector semiotic model (adapted from Iriskhanova & Cienki, 2018, p. 31)

To this day, the multi-vector model is the only model which tries to capture the complex nature of gestures based on their functional repertoire; however, since its proposal, it has not been used and elaborated in any of the existing research on gesture. It is still crucial to develop objective criteria for defining the exact place of a gesture on each vector and outline the ways to apply this model to the analysis of gesture.

The second approach was developed by Dominique Boutet (Boutet et al., 2021) in order to establish the link between the form of gestures and their function in a more objective way by tapping into the physical and anatomical components of movement. Using motion capture technology, Boutet and his colleagues focused on defining properties of gestures based on the variations of quality of movement, positions in space, and dynamics of each gesture articulation segment (i.e. shoulder, arm, forearm, wrist, phalanges, etc.). These properties were then mapped onto the semantics of the co-speech gesture to find regularities. For example, this analysis was used to establish connections between movement types and aspects in French. Although similar findings can be achieved through careful observation, kinesiological analysis undoubtedly provides a more comprehensive framework for these types of analysis. As has been mentioned in the previous chapter, kinesiological framework can offer unambiguous evidence for the existence and explain the nature of such concepts as recurrent gesture and gesture families. The works attempting to deploy kinesiological approach, for instance, explore relations between various linguistic notions withing grammar (e.g. negation (Boutet et al., 2021)) or facilitate cross-linguistic comparison between sign languages and gesturing (Chevrefils et al., 2021).

Although both multi-vector semiotic model of gesture and kinesiological approach provide appealing models of analysis, the classification used in the following chapters in the present paper hinges on more traditional ideas of McNeil (2005) and Kendon (2004) as adapted by Enfield (see Figure 4.4. from Enfield, 2013, p. 700) together with a dual distinction between the ways of gesturing proposed by Cienki (2013, p. 185). As Cienki puts it, the two main functions of co-speech gestures are pointing and representing. The basis of this categorization lies in gesture forms, which has proved to be a more solid ground for interpreting gesture meaning. And, if we look at the detailed explanation he provides for this distinction, it becomes apparent that these two functions can be logically mapped onto the classic gesture taxonomy and semiotic functions. Thus, for the purposes of the present work, the typology of gesture has been developed into the one shown in Fig 4.5.

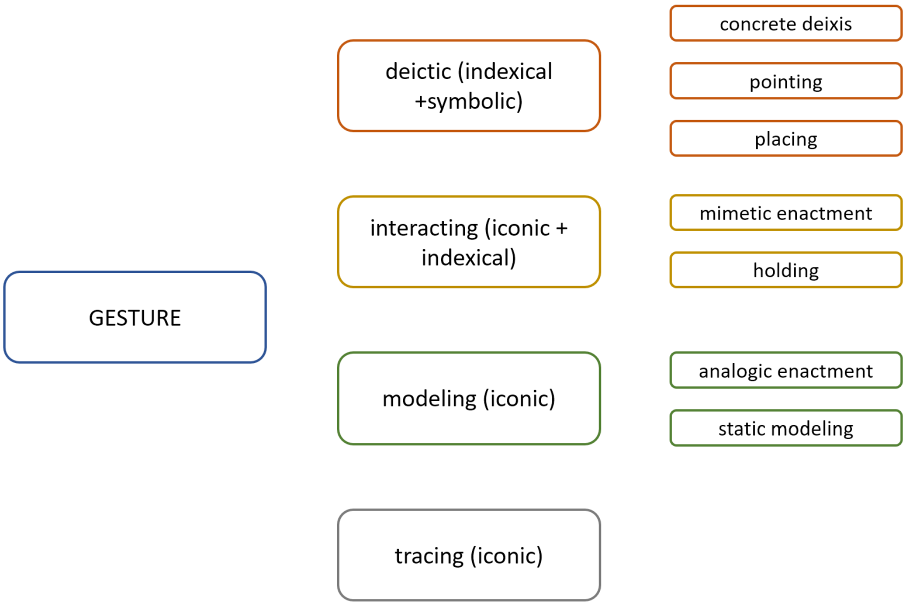


Figure 4.4. Gesture typology (adapted from Enfield, 2013, p. 700)

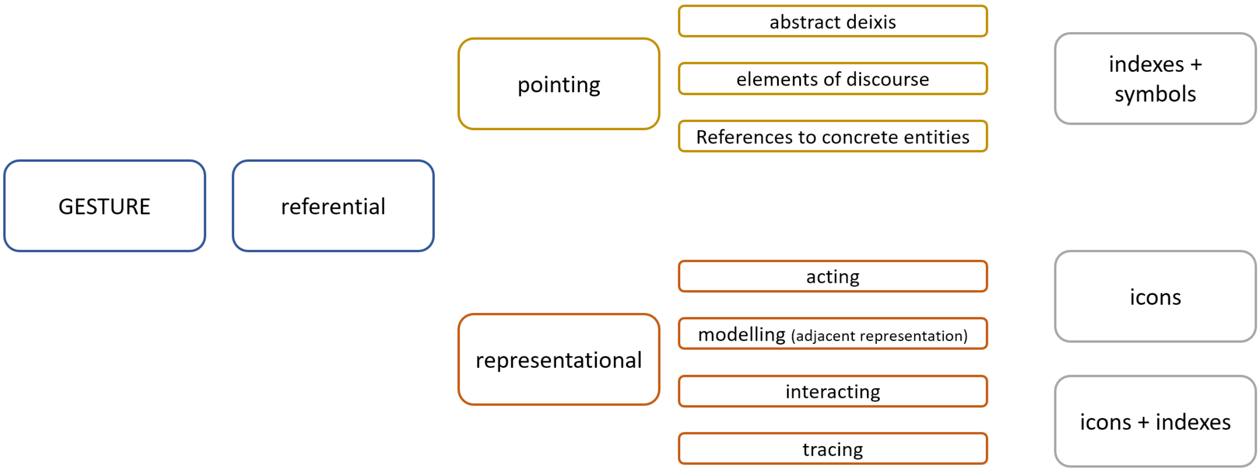


Figure 4.5. Gesture typology

## 4.2 Schemas

Schematization is an important process driving the semantics of gesture. Schemas in both verbal and non-verbal communication facilitate conceptualization, and retention of information, and enhance recognition of physical aspects of the referents (Cienki, 2013). There are two approaches to the way schematization works. First concerns the notion of image schemas (Lakoff & Johnson, 2003) which are defined as “basic patterns in our physical experience which provide the basis for understanding more abstract domains” (Cienki, 2013, p. 189). The second approach uses the term ‘mimetic schemas’ (Zlatev, 2007, 2014; Zlatev et al., 2005). It concerns exclusively the nature of gesture and is oriented towards more highly specified motoric actions which humans perform at various levels of abstraction. The supporters of the schematization approach also argue that it is also tightly connected with the notion of subjectification, the abstraction "away from the original instrumental function of such gestures as to be primarily discourse-functional, or pragmatic in nature" (Cienki, 2013, p. 190) in order to reflect speaker's personal view. This is claimed to be a diachronic process of pragmatic strengthening natural for both verbal and non-verbal communication. The linguistic view on image schemas and mimetic schemas also closely reflects the schematic and componential nature of perceptual symbols described in chapter 2 (Conceptual Structure and Spatial Structure).

### 4.2.1 Image schemas

In 1987, Johnson and Lakoff each published a book where they introduce a selected number of recurrent patterns (image schemas) derived from our embodied experiences with the world, such as PATH, CONTAINER, CYCLE, SURFACE etc., which shape our understanding of both concrete and abstract concepts (Johnson, 1987; Lakoff, 1987). The notion of image schemas is based on the idea that, instead of being static, propositional, and sentential, our knowledge is grounded in various dynamic patterns that emerge from our sensorimotor activities, such as "perceptual interactions, bodily actions, and manipulations of objects" (Lakoff & Johnson, 2003; Talmy, 1988, as cited in Gibbs, 2008, p. 239). Some of the image schemas are motivated by the iconic representation of referents that represent physical objects, others, - by metaphorical connections between a more abstract notion and an item with concrete reference. In other words, image schemas reflect spatial relations and movement patterns between the entities without reference to a specific viewpoint. These patterns are claimed to be all-pervasive and possess an internal structure which, by means of metaphorical extension, enables our understanding of abstract domains (e.g. grammatical forms in language (R. W. Langacker, 2014)).  In other words, image schemas are claimed to function as a source domain in metaphoric expressions. Such transfers become possible because image schemas motivate aspects of our thinking and reasoning, and they allow us to establish structural relations between seemingly non-related domains. This also provides a possible explanation for the nature of polysemous words; similar to that of different abstract domains, different meanings of a single word are motivated by our recurrent sensorimotor experiences with the world. Importantly, as image schemas are highly abstract representations of dynamic spatial patterns, which employ our experiences from different modalities (visual, auditory, kinesthetic, and tactile), they are argued to "operate on many levels of meaning construal, including grammatical and semantic structures" (Ladewig, 2020, p. 18).

Image schemas are widely used in semantic analysis; however, they generate particular interest in gesture research, since gestures, due to their nature, can naturally represent image schemas either as static entities or dynamic processes. Moreover, gestures can function as an interface between image schemas and spoken language, as they "can both reflect and influence the imagery content of an idea unit as it is being "unpacked" during speech" (Kita, 2000; McNeill, 2005, as cited in Cienki, 2005a, p. 422). Also, image schemas have a potential to explain metonymical and iconic nature of gestures and refute the idea of idiosyncratic nature of spontaneous gestures. This supports the idea that image schemas do not only semantically structure verbal expression, but also mediate the structure of gestural elements. When coupled with co-speech gestures, image schemas provide structural motivation for their form used by the speaker and facilitate their interpretation by the receiver. In other words, image schemas project conceptual structure onto gestural space and contribute to the appearance of recurrent gestures. Interestingly, as an experiment carried out by Alan Cienki and his colleagues (2005a) has shown, these two instances of use of image schemas can complement each other; gestures can complement information expressed in speech by making use of additional image schemas. For instance, in the experiment, the utterance "So in general on a regular test" which makes use of image schemas CONTAINER and CYCLE is supported by gesture representing image schema SURFACE (for details see Cienki, 2005a). Another experiment by Cienki (2005b) focused on the ways the recipients interpret different gestures. They showed that regardless of the function of co-speech gestures (referential (with abstract referent) or another), and the presence of corresponding speech (with speech or silent), participants could reliably attribute certain image schemas to the observed gestures. Later part of an experiment with concrete referential gestures, however, demonstrated recipients' increased reliance on speech to differentiate between the PATH and OBJECT schemas.

### 4.2.2 Mimetic schemas

While image schemas comprise abstract representations of embodied experiences, the theory of mimetic schemas accounts for a more concrete way of iconic representation of events, actions, or objects through the process of imitation, such as EAT, KISS, CUT, FALL, KICK etc. The concept of mimetic schema, coined by Zlatev and his colleagues, Persson, and Gärdenfors (2004), is grounded in the theory of developmental psychology and Piaget's idea that "the first cognitive representations in childhood arise through acts of sensory-motor imitation which are eventually internalized" (Zlatev, 2014, p. 4), as well as the theory of bodily mimesis. According to Zlatev, image schemas are "fairly specific, cross-modal, consciously accessible representations based on imitation, and largely shared within a (sub)culture" (Zlatev, 2007, p. 131). Moreover, as Zlatev puts it, they provide a basis for both verbal language and co-speech gesture, which helps explain their alignment. Their main property is that "they can help explain, almost literally, the "grounding" of *both* communication and thought through action and imitation, in both evolution and development" (Zlatev, 2014, p. 5: italics in the original). Thus, in co-speech gesture, mimetic schemas can be manifested in a form of primarily iconic representational gestures, in particular, acting and interacting, with different degrees of explicitness.

Additionally, mimetic schemas are argued to play a role in so-called discourse cohesive gestures: the type of recurrent gestures that allow for establishing connections between the parts of the story spread across discourse-time. In other words, they reflect the "embodied memory by the speaker in the sense of coming back to what you were doing before when you mentioned the same topic" (Cienki, 2017, p. 64). This use of the recurrent gestures is out of the scope of this paper; however, it opens up an interesting area of how co-speech bodily movements manifest themselves in discourse.

### 4.2.3 Comparison of the two theories and discussion

While both image and mimetic schemas emerge from our embodied experiences as "experiential structures not derivable from their composite parts" (Zlatev, 2014, p. 22), in comparison to abstract image schemas, mimetic schemas are much more specific and concrete representations of recurrent motor actions, such as EAT, SIT, TAKE OUT, etc. As Ladewig (2020) points out, mimetic schemas are more dynamic and can convey semantically richer information more closely characterizing the content of the speech. This naturally poses a question regarding the role of both types of schemas in gesture studies. Following Cienki's proposition (2013), Zlatev (2014) suggests that both image and mimetic schemas play a complementary role in the use of spontaneous gestures; however, they enter humans' repertoire at different stages of language development and represent concepts with different degree of abstraction. Mimetic schemas seem to play an important part in the ontogenetic development of gestures: gestures based on mimetic schemas are the first to enter expressive repertoire of a child (the idea similar to the distinction between two types of iconic signs mentioned in Section 1.1.4), while the role of image schemas in this process is less clear. Supposedly, image schemas emerge in children later, at around 3-4 years of age, as at this time occurs the "transition in the character of children's gesturing in the direction of greater abstractness" (Zlatev, 2014, p. 25). Hence, in the gestural repertoire of an adult, mimetic schemas are argued to remain the basis for pantomimic gestures, and representations of more abstract concepts correspond to image schemas. This observation can be also partially useful to draw inferences regarding the evolution of language and its bi/multi-modal origin.

In spite of their complementary function in application to iconic representational gestures, neither theory provides a clear account of the nature of pointing referential gestures with indexical ground. As Zlatev (2014) mentions, "imitation (…) appears crucial for the development of all three categories of gestures" but "*what is imitated is not of the same kind*" (Zlatev, 2014, p. 23, italics in the original). It is impossible to say that these two theories do not reflect spatial relations: they consistently capture both distance and motion as well as the guidance of action towards objects in space. However, they fail to account for relative position and concrete location of physical objects in space. In other words, they represent both "what" and "how", but not "where", if we approach the issue from the dual-pathway model of vision mentioned in (Section 2.2.1).

To sum up, both theories of schemas (image and mimetic) perform several important functions in linguistic research. Firstly, they provide explanation of how our embodied experiences can account for the way humans conceptualize their knowledge of the world, and, crucially, how this knowledge is represented in language. They also give insight into metonymic and metaphoric nature of both language and co-speech gesture. Moreover, schemas shed light into the systematic properties of gestures as more than idiosyncratic instances. They can also account for the existence of recurrent gestures. Additionally, the existing research on the role of image schemas and mimetic schemas in early language development, raises two fundamental questions regarding the universality of these phenomena across languages and cultures and the role of gestures in the evolution of language.

## 4.3 Metaphor and metonymy

Mittelberg suggests that "multimodal discourse is about abstract concepts and categories" (Mittelberg, 2013, p. 772). Thus, the formation of any sign relies on the process of abstraction, which, in turn, is driven by the mechanisms of iconic and indexical grounding. The form-meaning relationship in referential gestures is widely assumed to be guided by metaphorical conceptualization and/or metonymic extension (Cienki, Müller, Mittelberg inter alia). Conceptual metaphor explains the way abstract notions can be encoded in gesture using physical source domain. Conceptual metonymy, in turn, defines how gestures specify referents by partial representations of their contextually relevant features. It is claimed to lie in the core of all gesture types because gestures always only partially represent the referents (Mittelberg, 2013; Müller, 1998; Taub, 2001). Metonymy is argued to be the first step in creating metaphorical expressions: as Mittelberg and Waugh (2009, p. 329) put it, metonymy comes first in a "dynamic two-step interpretative model suggesting that metonymy leads the way into metaphor". Thus, these two concepts are not mutually exclusive; and the process of meaning construal can either depend on one of the processes at once or, most often, result from an interplay between both of them.

### 4.3.1 Metaphor

One of the first works that mentioned the role of metaphors in the process of gesture creation was Wilhelm Wundt's first volume of *Völkerpsychologie. Eine Untersuchung der Entwicklungsgesetze von Sprache, Mythos und Sitte* (Cultural Psychology. An investigation into developmental laws of language, myth, and conduct) (1900). Wundt developed a theory of language evolution out of expressive movements, claiming that gestures lie at the core of any symbolization process. He coined the term "symbolic gestures" implicitly considering the notions of metonymy and metaphor as a basis for such gestures. The class of symbolic gestures, according to Wundt, consists of "primary" and "secondary" symbolic gestures. In his interpretation, the meaning of primary symbolic gestures is non-transparent and arbitrary from the start, while secondary symbolic gestures undergo a two-step process from iconic gestures with traceable referents as an intermediate step to symbols "through a shift from (rather direct) iconicity to associations and indirect reference" (Teßendorf, 2014, p. 1545). Thus, for conventionalized symbolic gestures with consolidated form-meaning pairings, the iconic relationship between the base and the referent is practically lost due to constant use.  This way, Wundt's theory considers two possibilities for gestures to reify abstract concepts and pushes through the idea that iconic relationships together with metaphoricity underlie the creation of various types of gestures.

The interest in metaphorical nature of gestures emerged again only in late 1990's with the works of David McNeil (1985) and McNeil and Levy (1982). Most importantly, McNeil and Levy were also one of the first researchers to analyze metaphor in gesture with respect to CMT (Conceptual/Cognitive Metaphor Theory). This view became widely accepted in the field of gesture research, and the idea has been pursued further by researchers such as Cienki (2009, 2013, 2017), Müller (2013), Sweetser (1998), inter alia. The CMT claims that metaphors can be expressed in various forms of human behavior. The idea gained traction after the publication of Lakoff and Johnson's book "Metaphors we live by" (Lakoff & Johnson, 2003) where they defined the main features of metaphor as a tool for structuring, restructuring, and creating reality; however, it is important to note that the current CMT differs to some degree from the one originally introduced. The key tenets of the current CMT go as follows. Firstly, the use of metaphors is not restricted to literary style; on the contrary, metaphors are ubiquitous and are used in a range of language styles and registers. Thus, metaphors are considered to be part of a language user's mental lexicon with the vast number of cases of polysemy and idiomaticity seen as evidence for such an assumption. Secondly, conceptual metaphors are not random; rather, they are based on systematic mappings between conceptual domains. Importantly, in these pairings, there is a fairly stable correlation between abstract and concrete experiences, i.e., the source domain tends to use concrete notions while the target domain represents abstract notions. This was also reflected in the term "primary metaphor" coined by Joseph Grady (1997). Some examples of such transfers are IDEAS are OBJECTS, TIME is MOTION, GOOD is UP, KNOWING is SEEING, CAUSES are PHYSICAL FORCES, etc. As can be seen, unlike those metaphors that are grounded in the similarity between source and target domain, the conceptual metaphors are more likely to be based on our sensory-motor experiences with reality, which predicts that gestures should have an inherent tendency to convey concepts and ideas metaphorically. In fact, Cienki (2017) argues that many of the transfers from source to target domains "can be found in gestures rather than verbally in conversations (and other kinds of talk), suggesting that (…) metaphors play a fundamental role in terms of how we are thinking for speaking" (Cienki, 2017, p. 105). This account seems plausible for the majority of the source domains in CMT are grounded in the physical world. For instance, if we look at the list of conceptual metaphors stored on the server maintained by the University of California, Berkeley/metaphor@cogsci.berkeley.edu (Lakoff, 1994), we can see that out of 204 conceptual metaphors, 85% are drawn from our embodied experiences with the world, while only 15% transfer concepts from one abstract domain into another abstract domain (Figure 4.7.). All the 'embodied' domains can be, in turn, divided into 7 categories: physical entities (68), spatial (34), abstract (31), motion (21), proprioception (15), force (15), beings (10), activities (9) (Figure 4.6.). It is, thus, possible to assume that some of the conceptual metaphors, such as the ones with the source domains representing physical entities, spatial relations, motion, and force, can be naturally expressed in gesture and are, potentially, more often used in gesture or in gesture and speech together rather than speech alone. However, this suggestion is to be empirically tested.

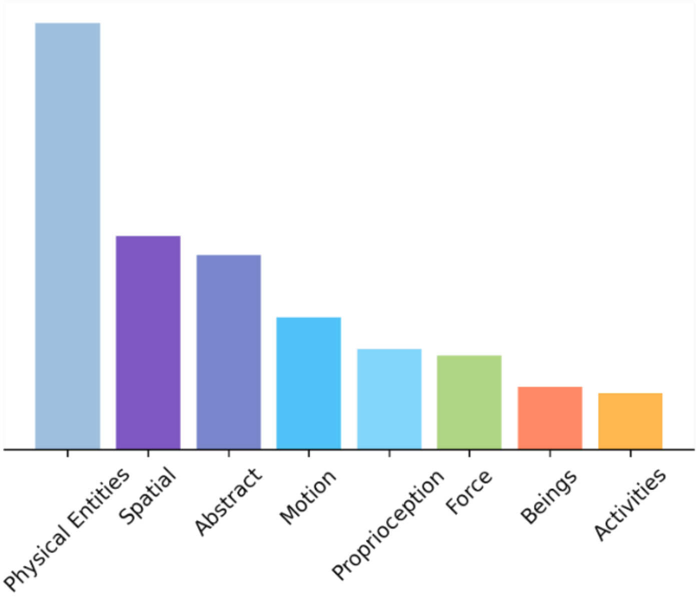


Figure 4.6. Distribution of domains in CMT

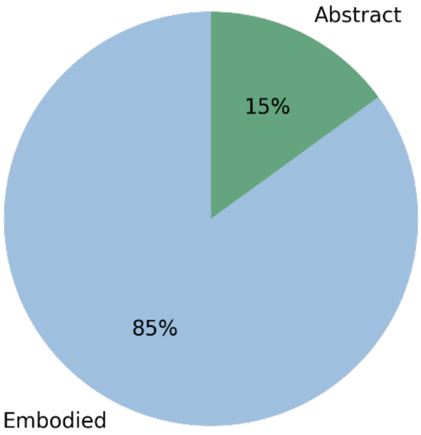


Figure 4.7. Distribution of abstract vs embodied domains in CMT

At the dawn of co-speech gesture research, McNeil (1992) distinguished several gesture types, among which there were iconic, metaphoric, deictic, beat gestures, etc. As can be seen, by defining a separate class of metaphoric gestures that he described as the ones "in which pictorial content presents an abstract idea rather than a concrete object or event" (McNeill, 1992, pp. 12–18), McNeil emphasized the role of metaphor for the process of semantic encoding and decoding in gestures. In fact, both iconic, deictic, and beat gestures are also often of metaphorical nature, which means that these categories and metaphorical gestures are not distinct types. As Cienki (2009, 2013, 2017) suggests, metaphor works differently for different gesture types as well as in different instances of their use; thus, it is useful to see metaphor as a scale or continuum, along which "the source domains may be represented in gesture in a more detailed way in (abstract) referential gestures, but then in more schematic ways as we move to the categories of discourse-structuring, pragmatic, and then interactional gestures" (Cienki, 2017, p. 100). This means that there is not an opposition between non-metaphorical representational gestures and gestures with metaphoric meanings, but rather a scale, one end of which is characterized by the highest degree of detail and elaboration and the other by the highest degree of abstraction with many instances in between.

The way the metaphor acts in representational gestures with iconic grounding is more obvious. We use all kinds of representational gestures to talk/gesture about various abstract concepts: describing concepts such as syntactic movement of constituents holding hands in a way that resembles holding a small tangible object (interacting), taking about ups and downs in a relationship by manually tracing the line as if shown as a graph in a coordinate system, where x axis represents the scale between bad (down) and good (up) and y axis stands for the time, etc. (for details Cienki, 2009, pp. 352–353). Pointing presents a more subtle case of how metaphor functions in gestures. Speakers of Indo-European languages typically employ two strategies when it comes to metaphorical pointing. The first is pointing at an abstract idea as if it is located somewhere in space, i.e. abstract deixis, thus reflecting IDEAS are OBJECTS conceptual metaphor. Sometimes, however, "metaphorically pointing to an idea can also be grounded, via metonymy, to a physical entity with which the idea is associated" (Cienki, 2009, p. 351). An example can be pointing at a person while referring to the idea or thought this person mentioned before. The most unclear relation between gesture and metaphor concerns beat gestures. Some of the researchers tend to exclude this category due to its close link to speech and speech prosody. However, they also seem important for the structure of discourse. An example from Casasanto (2008) suggests that the location in gestural space reflects metaphorical nature of beats: when talking about hot temperatures, speakers use beat gestures in the space above resting position, while mentioning lower prices for the car they want to buy, speakers use space below gesture position. Thus, apart from marking emphasis, beats also embody conceptual metaphors such as MORE is UP and LESS is DOWN in the example above. These complex relations can be also seen, in McNeil's (2005) terms, as two types of metaphors present in both speech and gestures: "expected" and "unexpected". "Expected" metaphors hinge on conceptual metaphors embodied in culture, making their form and content relevantly predictable. This type would encompass all the instances with the use of conceptual metaphor. "Unexpected" metaphors depend on the context and "form a bridge between the core idea unit or growth point of the utterance at the moment of speaking, and the larger discourse framework" (McNeill, 2008, p. 187). This category would reflect metaphorical basis of gestures beyond CMT, such as in case of interactional, pragmatic, and discourse-related gestures.

When it comes to the use of metaphor in verbal and bodily modalities, the patterns of alignment of speech and gesture might vary. Cienki (2009, 2013) describes three general trends: simultaneous production of metaphor in both speech and gesture, the use of metaphor in speech alone, and, lastly, the use of metaphor in gesture not being accompanied by metaphor in speech.

In the case of simultaneous production, it does not always mean the exact temporal co-occurrence of both modes: it can also mean partial temporal overlap or even consecutive use of gesture and speech. Interestingly, it is more typical for the co-speech gesture to be used cataphorically, i.e., preceding the use of verbal metaphor. Moreover, it might be natural to assume that metaphor expressed in speech is primary and the one produced manually is just its direct embodied reflection, which would make gestural and verbal metaphors produced together redundant. It indeed happens especially in cases when verbal metaphors contain reference to spatial source domain, but even then gestures tend to highlight different aspects of the source domain compared to speech.  For example, when a speaker defines 'honesty' as having three levels, they produce gesture with both hands held horizontally palm up and palm down delineating some area in their gestural space down to up respectively, potentially interpreting "three levels" metaphor as indicating the range of features these levels represent (Cienki, 2009). Nevertheless, speech and gesture often do not use the same source domain to refer to the target concept for the reasons that go as follows.  Firstly, as has been mentioned, referential gestures being a spatial medium refer more naturally to the entities located in real world and specify their location or other spatial characteristics. Thus, given the dynamic, spatio-motoric nature of gestures, certain source domains are more likely to be used by bodily rather than verbal modality (Cienki, 2017). In other words, although it is common for both metaphorical extensions to be aimed at expressing related meanings, they do not share the same source domain, and their relation should be rather seen as a complementary with different aspects of meaning are highlighted in speech and gestures. For instance, in a situation described by Cienki (Ibid.) the speaker was talking about moral values of being "right or wrong" and "black and white" substituting her speech with a chopping gesture "dividing" space into two parts. As Cienki explains it, the speaker supports verbal 'color' metaphor with the gestural spatial metaphor, "assigning two opposing 'colors' to the two spatial locations to make an absolute distinction between them" (Cienki, 2017, p. 98). Secondly, the use of metaphor in gesture in the absence of corresponding metaphor in verbal language would concern depiction of the abstract notion of time or events sequencing. As Cienki (2009, p. 360) puts it, "[p]eople do not always verbalize this metaphorical objectification", which, in turn, is frequently portrayed in gesture. For example, while for a large number of speakers of the languages whose written system is left-to-right oriented, the way to convey the notions of some action preceding the other consists of moving one or both hands left to right, and it is very unlikely that they will verbally express a sequence of events in spatial terms.

Although there has been a lot of research on the effect of metaphorical extension on gesture production, the filed still faces a number of difficulties researchers fail to overcome. Probably the greatest problem consists in the fact that the source domains encoded by CMT are not exactly the same in gesture: CMT metalanguage predominantly uses verbal mode, while gesture source domains are spatial and dynamic by nature. Thus, it poses a challenge for characterization of source domains which exist only in bodily modality. As Cienki (2005b, 2005a, 2017) claims, verbal formula is often insufficient. He gives an example of an American politician using PUOH gesture in order to represent the idea of supporting. The current approach would analyze this instance as a conceptual metaphor IDEAS are OBJECTS; however, as Cienki (2017, p. 100) puts it, using graphic or schematic image would be more representative, because "if [source domains] are spatial forms and dynamic forms, we inevitably lose something if we put then into words". For the example above, he suggests gestural conceptual metaphor SUPPORT is ‿, representing by this pictogram the form of the hand. Moving this idea forward, it is important to acknowledge the distinction between "metaphorical expressions on the level of words or gestures from metaphor as mapping on the conceptual level" (Cienki, 2017, p. 105): the way humans metaphorically use and interpret language and gesture reflect different aspects of our metaphoric thinking. Thus, the existence of metaphor in gesture and speech reflects the fact that human though processes are metaphorical; however, it also shows that it receives expression through various forms of behaviors. Although we can think about gestures as one of the reflections of how humans conceptualize the world alongside verbal expressions, using metalanguage adopted for analyzing verbal modality may give inadequate results because the experiences in which bodily domain is grounded comprise a partially autonomous system of representations. Hence, in order to analyze bodily modality as a system in its own right, new ways should be sought.

Another caveat applies not only to the analysis of metaphor, but to gesture research in general. A lot of evidence researchers have for the instances of co-speech gesture use comes from English and from some other languages in Europe and North America (German, Italian, Russian, Spanish). This means that the metaphorical mappings that prevail outside the Western world might differ, as is seen in the examples of perception of time (e.g. speakers seeing future as something behind them) and space (e.g. speakers using allocentric spatial reference system). This creates the need for cross-cultural analysis of language and gesture that would uncover patterns of mental imagery non-existent in the cultures outside Europe and expressed either verbally, gesturally or by the interplay of these two modalities.

To sum up, with respect to gestures, CMT has a potential to explain how the array of abstract concepts can be represented in gestures. As has been seen, certain types of metaphor in both speech and gesture are constructed in a similar fashion: by the process of mapping the concepts from a source domain on a target domain or, in other words, by thinking about one conceptual domain in terms of another; however, gestures due to their spatial and dynamic nature can uncover patterns that are never expressed verbally. Nevertheless, assuming that gestures constitute one of the means of expression of humans' conceptual system and make use of conceptual metaphors similar to natural language, makes it plausible to predict that gestural forms should have a certain degree of consistency. For example, metaphorical extension is likely to serve as one of the ways to conventionalize recurrent gestures and can be potentially extended to the interpretation of the majority of bodily movements speakers use.

### 4.3.2 Metonymy

Metonymy (technically speaking, 'synecdoche' or *pars pro toto* as we represent part of something which stands for the whole and not vice versa) is another way of conceptualization triggered by the process of using one concept in its direct relation to another. However, unlike in metaphorical relations, the source domain is salient and permanently present in metonymy since the transfer happens within a single domain. Moreover, in contrast to metaphor, this process largely differs in speech and gesture. In spoken language the use of metonymy is based on a "conventionalized system of form-meaning pairings" (Cienki, 2013, p. 188), while in co-speech gesture it is often grounded in the iconic relationship between the referred entity/action and the gestural form depicting part of this entity/action which is to be inferred by the receiver. Hence, metonymy can be seen as "the foundation of the construal of meaning in all gestures", because "gestures foreground only fragments of meaning" (Ladewig, 2020, p. 20).

Some of the gesture typologies, as described in Section 4.1. (Gesture classifications) reflect this inherently metonymic nature of representational iconic gestures by introducing different ways or modes used to isolate relevantly salient features of the referent with respect to the given context: acting, modelling, interacting, tracing (as shown in Figure 4.4.). For example, producing a "holding a pen and writing" brief motion can be used to convey the holistic idea of a "writing scene" or narrow the meaning down to the instrument (pen/pencil) depending on the context. In other words. the gesture shows only "parts of that form and the rest is inferred" (Cienki, 2013, p. 189).  Moreover, as has been seen in the analysis of the theory of PSS (chapter 2), selective attention is the primary mechanism which allows for the creation of schematic representations in a long-term memory around the situationally salient aspects of objects/events. This view is commensurate with Gärdenfors' (2014) view on visual processes which constrain semantic representations. Gärdenfors (as well as Ladewig (2020)) deems metonymy a primary semantic operation. He communicates an idea of attention shifts from one aspect of an entity to another which he calls *refocusing*. This mechanism is particularly pervasive and lies in the core of how we encode concepts using gesture and how we access semantic information conveyed via bodily modality.

Pointing gestures are also claimed to inherently entail metonymy. When it comes to pointing gestures with both concrete and abstract reference, "perceptually conspicuous site" is used "as the reference point for identifying the target" or "locatable index of the idea-as-space" (Cienki, 2013, p. 189). In other words, pointing gestures never cover the entire surface of the object, but rather to an area comparable with the tip of a finger or the hand in case pointing is signaled by a vertically oriented open palm. In the case of abstract reference, the metonymic extension is coupled with metaphorical conceptualization, since here an abstract idea of an invisible object is "reified in space"

(Cienki, 2013, p. 189): speakers often point at an empty space in their physical environment to either introduce a new idea or to refer to the one previously introduced in the course of interaction.

Following Jakobson (Jakobson & Halle, 1956), Mittelberg also uses the formal distinction between external and internal metonymy (Mittelberg, 2013). External metonymy uses relations that are spatial and pragmatic by nature, which means that there is no partiality in the relation between the sign and the referent, i.e. there is an outer contiguity between the sign and the referent. Internal metonymy, on the other hand, is based on the direct connection between one part and the other, a part and a whole, or vice versa. Importantly, internal metonymy "operates in all cases of gestural sign formation", while external metonymy "comes into play if the objects and figures depicted can be manipulated by the hands" (Ladewig, 2020, p. 20). This distinction contributes to the explanation of how abstraction works in the process of creating new signs. Externally metonymical gestures comprise signs with the indexical ground (pointing, location body index, hand-object index, hand-trace index) that take the human body as a starting point. Internally metonymical gestures have an iconic grounding (image, diagrammatic, and metaphor). The two poles are represented by image icon and metaphor icon, where icon metaphor, albeit similar to icon metaphor, requires additional semantic leaps in establishing similarity (Coulson, 2001). In other words, as Mittelberg explains it, icon metaphor exists only in one modality, for instance in gesture, but not in speech. Thus, it is possible to argue that both metonymy and metaphor are modality-independent (Cienki & Müller, 2008).

Importantly, as with metaphor, the representation of action in gesture possesses a certain degree of abstraction, i.e. "the handshape and movement are probably a schematized version (more relaxed) than would be used to actually perform the action" (Cienki, 2022, p. 5). Moreover, the parts of the reference depicted by gesture are also adapted for their use in bodily modality. This is true for both metaphor and metonymy in gesture: speakers use some kind of scaling in order to represent entities in gesture. Whether concrete or abstract, they are adapted in size to what is physically graspable and what can "fit" into the gestural space. This gives one more scrap of evidence that gestures often derive from instrumental actions (Streeck, 2009, 2013).

## 4.4 Conclusion: speech and gesture semantics

The chapter provided an overview of four cognitive linguistic notions (image schemas, mimetic schemas, metaphor, and metonymy) in their application to the use in co-speech gesture. While there is much similarity in how these phenomena can be represented and utilized in both spoken and manual(-visual) modalities, the number of differences based on the *affordances and limitations* of each modality can challenge and expand the understanding of these concepts. In particular, co-speech gesture displays high degrees of sensitivity to spatial characteristics of objects and events, which leads to *the representation of different aspects of humans' conceptual system* in a form that cannot be observed in the use of spoken language. Thus, the embodied memories find their natural reflection in the use of the gestures and gestural signs. This also leads to a different distribution of weight among the semiotic functions in both modalities, with an increased reliance on iconicity for bodily expressions. In sum, the chapter showed that although similar principles of establishing meaning lie in the core of different modalities, there is a number of *modality-specific features* that cannot be neglected.

# 5 Syntax

According to the main tenet of cognitive grammar formulated by Langacker (1986), "Grammatical structures do not constitute an autonomous formal system or level of representation: They are claimed instead to be inherently symbolic, providing for the structuring and conventional symbolization of conceptual content. Lexicon, morphology, and syntax form a continuum of symbolic units, divided only arbitrarily into separate 'components'" (R. W. Langacker, 1986, pp. 1–2). It is also assumed that language is tightly related to the human's experiences with surrounding physical reality which, in turn, serves as the driving force behind the processes of conceptualization responsible, among other things, for the choice of lexical and grammatical construction and, subsequently, for meaning construal. Another process crucial for meaning construal and the acquisition of lexical items and grammatical constructions is schematization, the "process of extracting the commonality inherent in multiple experiences to arrive at a conception representing a higher level of abstraction" (Langacker, 2008, p. 17). Both the processes of conceptualization and schematization are essential for the creation of schemas, that can be found on all linguistic levels, including gesture.

Importantly, as Ladewig (2020) emphasizes, "speech events are situated in an ongoing discourse" (Ladewig, 2020, p. 8). The interaction proves to be successful in the case of interlocutors' joint attention on the same conceptual entity, in the condition of the limited attentional and conceptual scope. Discourse is, in turn, multimodal, which means that gestures produced in parallel with speech can provide cues to the disambiguation of linguistic context and the shift of focus in online processing, thus complementing verbal information. Also, according to Langacker (2008), co-occurrent gestures are important for the process of conceptualization and "developing an integrated account of grammar, processing, and discourse" (Langacker, 2008, p. 249). They "seem to be motivated by embodied conceptual structures". This makes it possible to establish certain patterns of similarity between verbal language and gesture.

Contrary to the belief (McNeill, 1992, 2005) that gestures represent holistic units where distinct features are determined by the meaning as a whole, Kendon (1995, 2004) contends that the system of co-speech gesture is organized hierarchically. which means that single spontaneous gestures are constructed from clearly defined phases that, in turn, can be linearly combined into gesture phrases and further on into higher-order gesture units (Figure 5.1). Kendon (Kendon, 2004, p. 111) identifies gesture unit as an entire "movement excursion" between the positions of full relaxation. The movement of the hand during this "excursion" is not uniform and can be defined by a number of distinct phases when the quality of the gesturing changes. Thus, Kendon breaks gesture unit into four phases:  preparation, pre-stoke hold, stroke, post-stroke hold and recovery. At the same time the first two phases with the exception of recovery comprise a gesture phrase. This distinction between smaller and bigger units was emphasized in order to facilitate the analysis of series of movements. Moreover, within gesture phrases the most important meaning bearing element is a stroke which is defined as the "phase of the excursion in which the movement dynamics of 'effort' and 'shape' are manifested with greatest clarity" (Kendon, 2004, p. 112).

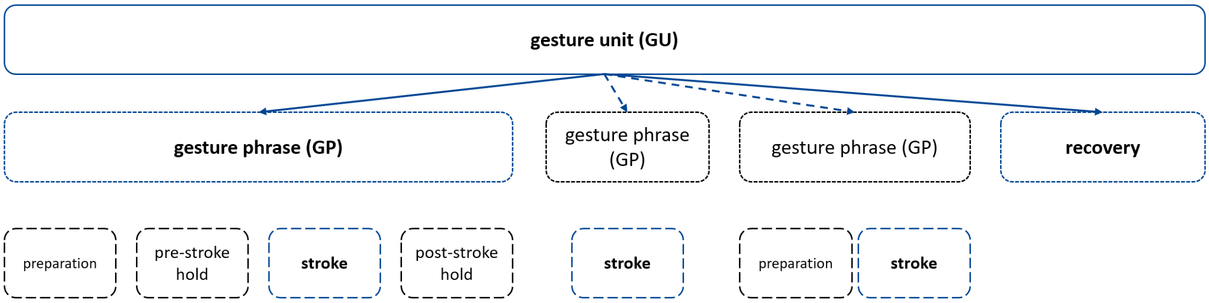


Figure 5.1. Linear structure of the gesture (after Kendon, 2004, pp. 111–113)

The ideas of Kendon recently received empirical and theoretical support in the research of Müller (2014, 2018), Ladewig (2014c, 2020), Bressem (2014), Cienki (2005b, 2013), Zlatev (2014), Pouw et al. (Pouw, de Wit, et al., 2021; Pouw, Dingemanse, et al., 2021) inter alia. They corroborated the claim that gestures have linguistic potential in their structural and formal features. They also introduced the notion of "compositionality of gestural form" and proposed that "gestural meaning may be composed of isolated features" (Ladewig, 2020, p. 16). Gesture research has gathered substantial evidence of compositionality of gesture forms contrary to the initial idea of gestures being holistic units (e.g. McNeil, 1992, 2005). The analysis of recurrent gestures and gesture families shows that gesture forms can be decomposed into smaller meaningful segments (e.g. kinesthemes) and, thus, possess at least rudimentary morphology. It has also been argued that since recurrent gestures display systematic behavior between the four structural parameters of gesture use (see Recurrent gestures and gesture families) and gesture meaning, they must be as well related on a conceptual level. As a result, a body of research concerned with recurrent gestures, repetitions in gesture, gesture families, gesture kinematics and image/mimetic-schemas was built.

Since gestures are treated as motivated signs, it is possible to predict that the analysis of their forms shed light on how combinations of gesture with speech and/or gestures with other gestures lead to the construction of new meaning. However, the question regarding gestural grammar pertains: do we speak of multimodal grammar where gestures are one of the inherent elements of composite utterances or do gestures themselves form a system of their own which interfaces with that of a spoken language? This question is difficult to answer because, in the current stage of development, *gestures are not put under pressure of creating their own system for communication due to their tight connection with well-developed system of spoken (or sign) language*. Fricke (2013, 2014b) proposes a distinction between two processes: code manifestation and code integration. She defines code manifestation as a capacity of language to manifest itself into different modalities (e.g. vocal, bodily). In turn, code integration means embedding of a modality (e.g. bodily) into a more dominant one (e.g. vocal), i.e. in case of code integration multimodality is achieved "within grammars of single languages on the level of the language system" (Fricke, 2013, p. 751). Code manifestation for the most part requires higher degree of conventionalization of gestures, while code integration is not limited to the level of conventionalization. This distinction also reflects two ways the syntactic properties of gestures can be identified. On the one hand, researchers talk about syntax in gesture from a form-based view based on the assumption that gestures are motivated signs. Thus, the way different gesture forms are coordinated is reflected in the meaning that is constructed as a result of this combinations. In other words, gestures are treated as constituents similar to syntactic units in vocal (or sign) language. On the other hand, syntax in speech-accompanying gestures can be approached from functional/structural perspective. In this case, gestures are analyzed together with speech and the main focus of such research is their functional integration into verbal syntax. Thus, according to this approach, the meaning of a gesture is shaped by the syntactic information given in speech regarding the syntactic slot it occupies.

## 5.1 Code integration

As has been seen in Chapter 4 (Semantics), gestures have a strong depictive capacity grounded in experiences of our bodies with the world, which enables them to convey meaning and stand in for linguistic units. As a result, gestures can be functionally integrated into the structure of verbal utterances. In many cases gesture produced simultaneously with speech foregrounds and/or complements the meaning expressed in the verbal utterance as shown in Figure 5.2. from Ladewig (2020). As a result, it is possible to talk about *multimodal gesture-speech ensembles*, where the gestures highlight conceptual archetypes of different grammatical classes to make certain aspects conceptually salient.

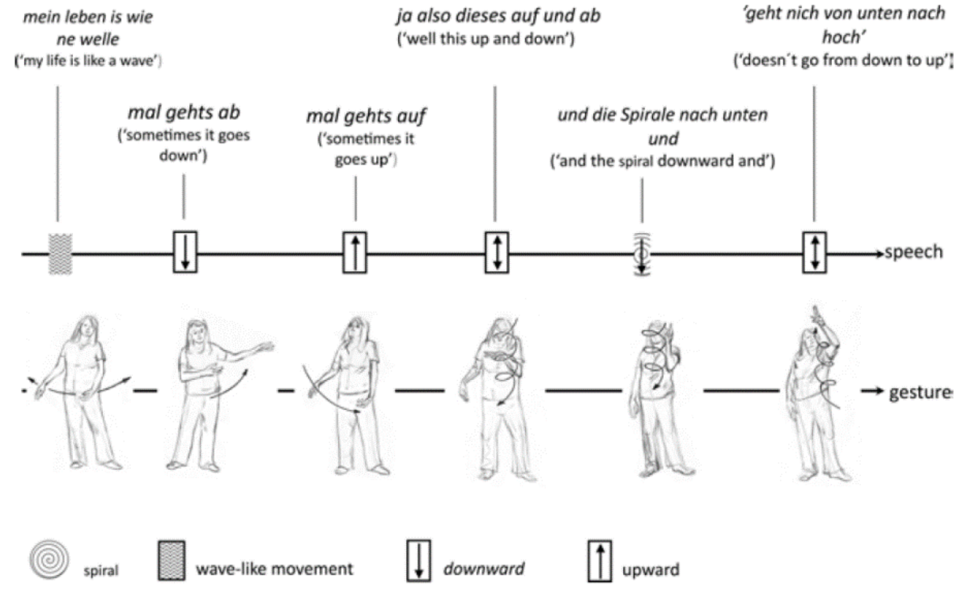

mein leben is wie ne welle
('my life is like a wave')
mal gehts ab
('sometimes it goes down')
mal gehts auf
('sometimes it goes up')
ja also dieses auf und ab
('well this up and down')
und die Spirale nach unten und
('and the spiral downward and')
'geht nich von unten nach hoch'
('doesn´t go from down to up')
speech
gesture
spiral
wave-like movement
downward
upward

Figure 5.2. Gestures foregrounding the verbal meaning (Ladewig, 2020, p. 95)

The process of structural integration is also tightly related to the created of multimodal constructions - conventionalized speech-gesture co-occurrences based on stable combinations introduced by 'gesture-attracting items' (Kok, 2016) or triggers. Four main functions of gestures as syntactic constituents have been analyzed: nouns, verbs, attributive adjectives modifying nouns, and adverbials modifying verbs.

Firstly, gestures possess an *ability to reflect conceptual archetypes of nouns and verbs.* Famous examples come from research by Ladewig (2013, 2020). If for any reason speakers cannot use the respective word for an entity they are talking about (e.g. they do not know what it is called or they are in a situation modelled as a Tabu game), they tend to insert gesture to supplement their verbal utterance. An illustration given in Ladewig (2014c) shows how a speaker attempts to explain the word ‘draisine’ (‘handcar’) by uttering ‘Ach hier mit diesen’ (‘Well, here with these’) and producing a multi-stroke gesture by moving their arms up and down with their fists clenched as if physically pulling and pushing some mechanism. Importantly, the interpretation of gesture is driven not only by the observable movement based on an action produced by driving a handcar, but also by the surrounding elements. In other words, “the syntactic position foregrounds semantic aspects of the gesture” (Ladewig, 2014c, p. 1672). In a given example, the demonstrative pronoun ‘these’, which acts as a gesture-attracting item, together with a gesture create a determiner phrase (DP), which on a higher level together with the preposition ‘with’ builds a multimodal prepositional phrase (PPmumod) (Figure 5.3.). Thus, when interpreting this gesture, recipient would not only focus on the reenactment of an action, but also on both syntactic and semantic structure of an utterance, which would help to disambiguate the meaning of a produced gesture as it encodes information about the object as a dynamic gestalt. This allows us to talk about the syntax-semantics interface where the motivation for gestural form is defined by one of the mechanisms as discussed in Section 3 (Semantics) which is mapped onto the identified semantic position to create a complect multimodal meaning. As explained from the cognitive grammar point of view, “speech can be considered to trigger and impose specific mental operations on the gestures” (Ladewig, 2020, p. 100). This phenomenon is also explainable through the processes of summary and sequential scanning introduced by Langaker (1986, 2008) which allow us to construe the whole scene either as an atemporal state or as a dynamic process. Again, similar processes are observed in human vision which enables us to gain a rapid global assessment of the scene or to access more fine-grained information by shifting the focus of attention from one aspect of the stimulus to another. Having taken this into the account, it is possible to suggest with the presence of speech, humans are able to reliably interpret gestural meaning not only by identifying the syntactic slot they occupy in the utterance, but also by focusing on the entirety of behavior of their interlocutor (for nouns) or on the selected aspects of this dynamic process (for verbs and attributes).

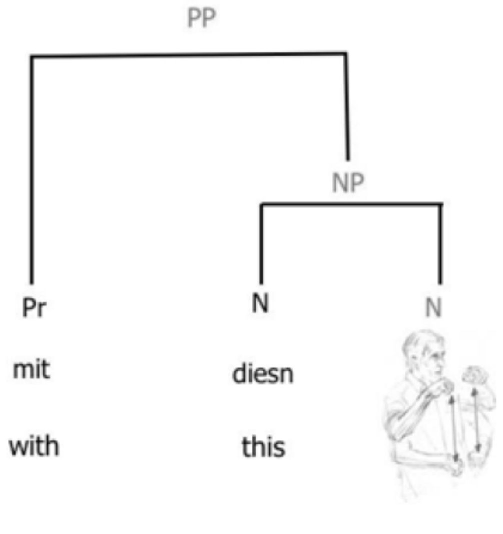


Figure 5.3. Syntactic structure of a PPmumod (Ladewig, 2014c, p. 1666)

Another example by Ladewig (Ibid.) and illustrated the use of gesture in a syntactic position of a verb in an utterance ‘Und wir hinten’ (‘And we from behind’) + pushing movement. In general, the execution of gesture together with the verbal utterance is similar to the one described above: dynamic movement performed together with speech is interpreted due to its integration into the syntactic structure of a sentence. Interestingly, however, unlike the first example, where gesture is integrated linearly, following the verbal expression, here the stroke temporarily co-occurs with uttering the second syllable of an adverb 'hinten'. Nevertheless, temporal alignment of gesture and speech does not play a significant role here; what is important is that they happen with close temporal proximity which allows for the interpretation as shown in Figure 5.4.

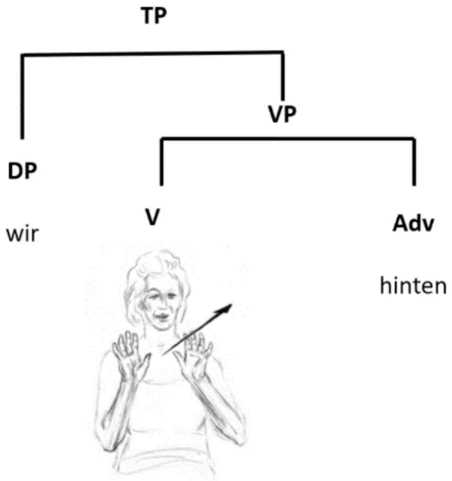


Figure 5.4. Syntactic structure of TPmumod (after Ladewig, 2014c, 2020)

As has been mentioned, most of the gestures occupying syntactic slots are representational gestures with different levels of conventionalization. However, as statistics shows (Ladewig, 2020, p. 80), spontaneous form-meaning occurrences are used in 63,3% of the cases when gestures occupy a slot of the noun and 68% when they take syntactic function of a verb, which means, that conventionalization is not required for gesture integration. When perceived without speech, such spontaneous bodily movements are reported to be more semantically loaded: the information communicated linearly by words can be transmitted simultaneously by gestures analyzing both the hand shape, space, orientation of the palm, and the movement the hands produce. In order to convey particular meaning spoken language is required to constrain their meaning. As has been seen in the example of a handcar, although the speaker refers to an object, the produced sign is dynamic and based on the action rather the static depiction of an item in question. Hence, the question regarding whether spontaneous gestures display morphological features of nouns and verbs by foregrounding certain parameters (gesture space, character of the movement, etc.) remains open.

Finally, gestures frequently function as adjectives and adverbs that modify verbal nouns and verbs respectively. This feature of gestures is referred to as "multimodal attribution", and it presents a great challenge to the researchers due to the variety of cases in which gestures are performing this function.

On the one hand, they can be used as a modifier that specifies form and size of an object (Bressem, 2014; Fricke, 2013, 2014b). For example, defining a type of bottle holder used in Italian restaurants, the speaker said, "in som Metallding drinne" ("in such a metal thing") and produced a gesture with both hands oriented downwards in an arced shap and moving left to right modifying "such a metal thing" with a depiction of its shape (e.g. "bent") (Figure 5.5.). It is also notable that the German article 'son' ('such') functions as a gesture-attracting item similar to pronoun "diesen" in one of the examples above.

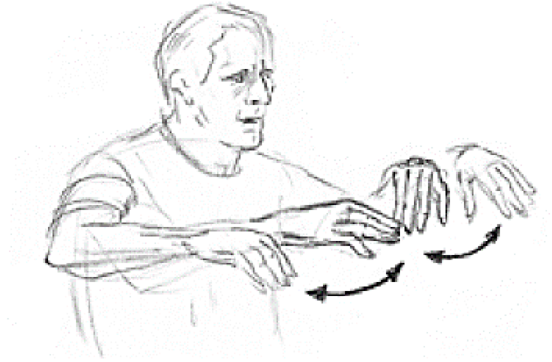


Figure 5.5. Gestural attribute (Bressem, 2014, p. 1643)

While the ways by which gestures can modify nouns are supposedly limited to the instances like the one above which are triggered by the gesture-attracting items in speech, the gestural modification of verbs is a more unclear case. Here, this will be illustrated by looking at the concept of manner. Due to their depictive power, gestures can provide information about the character of a process which may be absent in speech for various reasons. Classic example of this case is Kendon's (2004) observation of a farmer who is talking about the stages in cheese production. While talking about how, in order to prevent the cheese 'sweating', farmers scatter ground rice to absorb the moisture, the man uses a generalized verb "throw". However, while the process of throwing can be produced in any number of ways, the speaker uses a gesture by shaping his hand in a way showed in Figure 5.6. which enacts holding some loose material and scattering it with small swaying motions. As a result, gesture used by the speaker is much more semantically loaded and provides the listeners with a rich representation of the manner in which the action was performed.

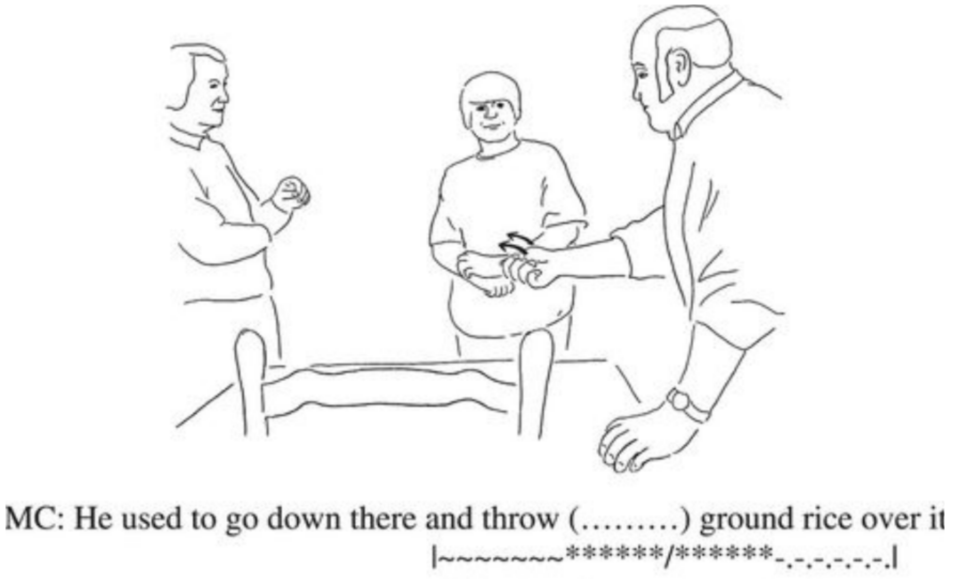

MC: He used to go down there and throw (.........) ground rice over it
|~~~~~~~~*******/*******-.-.-.-.-.|

Figure 5.6. Gesture performing a function of adverbial modifier (Kendon, 2004, p. 114)

Gestures also have an ability to encode spatial details in addition to the manner of motion, such as path and direction of motion, even if they are not expressed verbally. McNeil and Duncan (2000) showed that speaker of Spanish sometimes complement verbal expressions of path with the gesture encoding both path and manner. Similarly, one of the cross-linguistic experiments conducted by Kita and Özyürek (2003) demonstrated that the speakers of three structurally distinct languages, English, Turkish, and Japanese, systematically encode directional information (path) when talking about Swing events. More recent research on Czech language (Fibigerová & Guidetti, 2022) showed that the encoding of the path in gesture is more consistent, while the depiction of manner is less frequent and predominantly used by adult speakers. At the same time, if encoded, the character of movement is conveyed in gesture with more precision than in speech. Hence, it seems logical to argue that gestures performing a function of adverbial modifier reliably provide additional semantic information absent in speech. However, according to the statistics, the speakers, in fact, tend to gesturally express information that is already present in speech. In other words, gestures used with the verbs of motion are more likely to be constrained by the ways semantic and grammatical information is encoded linguistically. The samples used in all these studies are relevantly small; thus, the notion of multimodal attribution requires more principled research to reach definite conclusions.

All the examples given above support the idea of gestures being tightly integrated in the spoken/sign languages either by filling in certain syntactic slots which are defined by the syntactic structure of verbal utterances or by being used redundantly or co-expressively together with the word signaling the same or similar meaning. In either scenario, decoders still rely on syntactic information present in speech to infer the meaning, which means that the system of gesture interfaces with the syntactic structures of a spoken language rather than creating a system of grammar of their own.

## 5.2 Code manifestation

The grammatical potential of gestures lies in their capacity to create organized temporal sequences as has been shown in the example of gesture phrases and gesture units (Figure 5.1). This linear organization of gesture sequences can manifest itself systematically in three types of combinations: gestural repetition, recurrent gesture sequences, and gesture scenarios.

Firstly, Jana Bressem (2014) brings up a notion of gestural repetitions which she divides into two types: iteration and repetitions. Iterations are repetitions of gestures without significant changes in direction, quality and/or location in space in which "the successive execution of strokes does not result in a new and complex gestural meaning" (Bressem, 2014, p. 1644). Structurally, iterations consist of two gesture phrases with similar stroke phases. They perform discursive and meta-communicative functions by highlighting and complementing certain aspects in speech and resemble word repetition in speech. Many examples of iterations are also very similar to the examples discussed in the Section 5.1 above (Code integration) as they concern attributive function gestures. A more autonomous phenomenon of reduplication comprises two different gestures produced in close temporal proximity to each other which contribute together to the creation of a new complex gestural meaning. They are also produced together with speech and support the verbally expressed meaning, but, unlike iterations, become independent entities governed by their own structural rules not guided by accompanying speech. For instance, reduplications can express two notions: Aktionsart ('iterativity') and plurality. The case of iterativity is illustrated by an example where the speaker utters "send back and forth" while producing several movements with their hand away (the first stroke) and towards (the second stroke) the body, thus conceptualizing multiple repetitions of movement in space. The notion of plurality, in turn, can be exemplified by a situation where a speaker says "read through the single steps" while simultaneously producing three strokes with the curved palm each at different levels in gestural space. These strokes represent single steps the speaker is talking about, thus, together creating a complex grammatical meaning.

Gestural repetitions can be seen as an intermediary stage between the processes of code integration and manifestation. On the one hand, both iteration and reduplication show close connection to speech based on their temporal co-occurrence. The two common functions of iterations are prosodic/pragmatic and emphasizing, which reflects their dependence on spoken language. On the other hand, certain examples of repetition and reduplication in particular are driven by recurrent gestures. Moreover, the instances of reduplication tend to display higher degrees of conceptualization, abstraction, and grammaticalization. Interestingly, instances of reduplications in gesture used to convey the notions of plurality and iterativity show similarities with sign languages in structure and form-meaning mapping. This means that the complex meanings created by reduplications are less tightly connected to the semantics of verbal utterances.

Fricke (2013), however, develops these ideas further and insists on distinguishing between iteration or repetition and recursion in gesture. She (Ibid.) claims that not only do gestures have natural capability to form long flat structures with sequences of gesture phrases independent of each other, but they also possess properties that allow for self-embedding of gestural constituents. The main argument Fricke provides for her claim is that gesture units are delimited by the phases of complete relaxation and return into the rest position. As some of the video sequences analysed in Fricke's paper show, these phases are characterized by different degrees of relaxation, which allows researchers to argue that partial relaxation marks secondary gesture units embedded into larger ones as shown in Figure 5.7. This view is in keeping with research by Ladewig (2020), Müller et al. (2013), etc. on gesture scenarios.

$$
\begin{array}{lll}
\mathrm{GU} & \rightarrow & \left\{ \begin{array}{l} \mathrm{GP\ Retr} \\ \mathrm{GU\ GU(GU1\ldots GUn)\ Retr} \\ \mathrm{GU\ GP\ GP(GP1\ldots GPn)Retr} \\ \mathrm{GU\ (\ GU1\ldots GUn)\ GP(GP1\ldots GPn)\ \ (GUn+1\ldots GUz)\ Retr} \\ \mathrm{GP(GP1\ldots GPn)GU(GU1\ldots GUn)\ \ (GPn+1\ldots GPz)\ Retr} \end{array} \right\} \\
\mathrm{GP} & \rightarrow & \mathrm{(Prep)\ SP} \\
\mathrm{SP} & \rightarrow & \mathrm{S(\ S1\ldots Sn)} \\
\mathrm{S} & \rightarrow & \mathrm{(Hold)\ s\ (Hold)}
\end{array}
$$

Figure 5.7. Recursive structures in gestures (Fricke, 2014b, p. 744)

*Gesture scenarios*, according to Müller et al. (2013), are combinations of gestures depicting objects, actions and/or events into longer sequences. Two criteria for seeing scenarios as larger units of meaning are close temporal proximity characterized by the no return to complete rest position in between gesture units and common gestural space. The maintenance of the same gesture space is an important factor that also ensures formal coherence. This means, that the speaker creates a spatial frame in order to communicate the complex situation as is shown in Figure 5.8. The speaker gesturally delimits the scene by defining the upper sector of their gestural space ('blauen himmel'/'blue sky') as a relative spatial position, thus creating a stage for other events unfolding in the story. In other words, the meaning of gestures feeds on the meaning of preceding gestures.

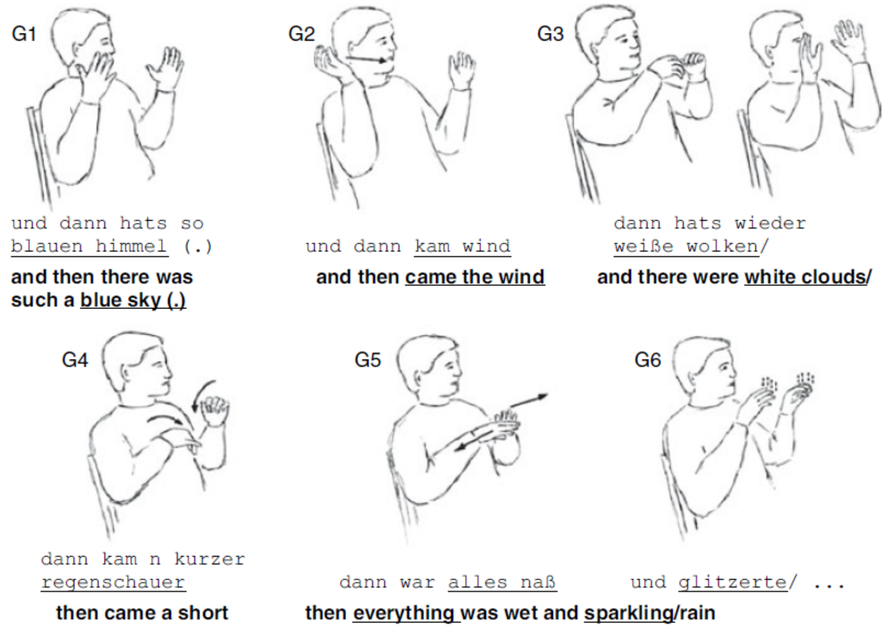
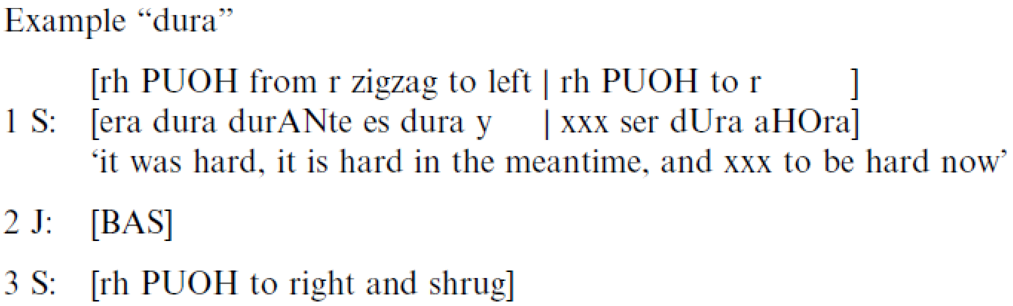

G1
und dann hats so blauen himmel (.)
and then there was such a blue sky (.)
G2
und dann kam wind
and then came the wind
G3
dann hats wieder weiße wolken/
and there were white clouds/
G4
dann kam n kurzer regenschauer
then came a short
G5
dann war alles naß
G6
und glitzerte/ ...
then everything was wet and sparkling/rain

Figure 5.8. Gesture scenario (Müller, Bressem, et al., 2013, p. 724)

The third type of structure is called gestural sequences (Ladewig, 2014c, 2020; Teßendorf, 2014; Seyfeddinipur, 2004). Although their structurally resemble gestural reduplications, the key feature of sequences is that they are composed from conventionalized or recurrent gestures. They tend to "operate on the thematic structure of an utterance, marking its topic and comment" (Ladewig, 2014c, p. 1566). Similar to gesture scenarios, gestural sequences are not only related to the simultaneously produced utterances, but also to their "gestural neighbors". One of the examples is the use of cyclic gesture together with PUOH (palm up open hand) gesture analyzed by Ladewig (2011). For instance, when used as a turn-holding device or a marker of verbal disfluency, cyclic gesture indicated the process of mental search of a concept concluded by the use of PUOH gesture to present the concept that has been retrieved. Alternatively, PUOH gesture can precede the production of cyclic gesture if the latter is used in a request: the idea is presented by PUOH and followed by cyclic gesture indicating a demand. Teßendorf (2014) also provides two examples of different combination of brushing aside gesture (BAS) with PUOH gesture:

a) a sequence of PUOH gestures followed by BAS to present a number of ideas and then dismiss them. The speaker is talking about a number of reasons (supported by PUOH gestures) why renting an unfurnished apartment for one month and paying a deposit of two months above that is not worth it concluding the enumeration by saying "no tenía absolutamente nada" ("there was absolutely nothing in it") and performing one BAS gesture;
b) a sequence of PUOH gestures and a BAS gesture produced by two interlocutors in a form of a dialogue without speech as shown in Figure 5.9. In this mini conversation (lines 2 and 3) conducted entirely in gesture, speaker J uses BAS gesture not only to indicate the lack of sympathy towards S, but also shows their disagreement about the topic. Speaker S. in turn, performs a PUOH gesture with a shrug meaning something like "well, let's agree to disagree", indicating that they quit the topic at this stage.



Example "dura"

```
      [rh PUOH from r zigzag to left | rh PUOH to r       ]
1 S:  [era dura durANte es dura y    | xxx ser dUra aHOra]
      'it was hard, it is hard in the meantime, and xxx to be hard now'

2 J:  [BAS]

3 S:  [rh PUOH to right and shrug]
```

Figure 5.9. Example of gesture sequence (Teßendorf , 2014, p. 1552)

To sum up, all the findings on linear constructions and compositionality in gestures show that they have a lot of parallels with spoken and sign languages. This allows to look at the idea of complete grammaticalization of gestures that is going to be discussed in the next section. At the same time, due to specific affordances of manual(-visual) modality, both spatial and temporal organization play equally important role in structural properties of co-speech gestures.

### 5.2.1 Grammaticalization of gestures

The process of grammaticalization has been extensively researched by cognitive scientists in diachronic view on language as a way to analyze how the reanalysis of form-meaning relations cause language change.

It has already been shown in the examples above (e.g., gestural sequences) that sometimes recurrent gesture patterns can act as monomodal constructions without the presence of speech. This corroborated the view that the from-meaning mapping in both spoken language and gesture follows similar conceptualization paths and depends on similar cognitive constraints. Moreover, following the idea of thinking for speaking, the notion of "thinking for gesturing" (Cienki & Müller, 2008) emerges. It reflects the idea that "a particular kind of thought is mobilized for the purpose of gestural communication, reflecting imagistic aspects of thinking" (Ladewig, 2020, p. 34). This led to a number of experiments that obtained evidence that certain form parameters in gestures can be stabilized in their form-meaning pairings, thus, revealing the symbolic status of gesture units. This stabilization of form and meaning is based on recurrent hand shapes and movements tied to particular pragmatic functions and/or used as grammatic markers. It has also been noted that a lot of similarities can be observed between gestures accompanying speech and sign languages which suggests that grammatical markers in sign languages evolved from gestures used together with spoken discourse (Ladewig, 2014c, Pfau & Steinbach, 2006). An example of this is the use of PUOH gesture as a discourse-pragmatic marker conveying agreement or seeking agreement also observed in American, Danish, and Turkish sign languages (Van Loon et al., 2014). Another example supporting the view of grammaticalization in gesture is the use of cyclic gesture as a functional marker of verbal aspectual meaning (Ladewig, 2020). It has been shown on the example of three languages, English, German, and Farsi, that irrespective of the stative or eventive status of the verbs in a spoken language, they are frequently accompanied by Cyclic gesture, which, according to Harrison & Ladewig (2021) is an element of multimodal progressive utterances which consist of be + present participle + cyclic gesture. In other words, even in the absence of progressive aspect in a language (e.g., German) or the choice of stative verb in any of the languages, speakers perform cyclic gestures to foreground the idea of the durativity and continuity of the action. One more argument in favor of possibility for schematization and decontextualization of cyclic gestures (and, potentially, other recurrent gestures) is that the movements depicting continuous rotation of a palm that constitute the family of cyclic gestures is used to mark aspect in sign languages (e.g., ASL, LIS).

Having developed interest in these regularities, Wilcox (2017) suggested two paths explaining how gestures can enter the linguistic system of sign languages. This model can also be extended to the use of gestures in monomodal constructions (both performed in space during the interaction or used in a form of emojis in text) by speakers from hearing communities who predominantly use spoken language as shown in Figure 5.10. According to the first path, gestures enter the system as lexical morphemes and through the process of grammaticalization starts functioning as a grammatical morpheme. Wilcox and Ruth-Hirrel (Ruth-Hirrel & Wilcox, 2018; Wilcox, 2017) provide a number of examples of this first route, both recent and going back centuries. One of them reflects the emergence of a grammatical morpheme 'can' from the gesture showing the muscular strength of the upper body which got lexicalized as a lexical morpheme 'strong' when it entered ASL. Nowadays, we can see the similar use of a version of this gesture as a part of spoken and written (emoji) utterances conveying the meaning "You can do it!" or "Brace yourself!" both together with speech and independently. Alternatively, gesture (or even a facial expression) can enter the language as a pragmatic marker or marker of prosody and, again, through the process of grammaticalization, turns into a grammatical morpheme. The example of this route is related to the use of PUOH gesture originated in the instrumental actions of showing or presenting an object. However, as Van Loon et al. (2014) point out, it is difficult to talk about full semantization of this gesture because it "did not lexicalize into a predicate that conveys a meaning like "suggest" or "put forward (an idea)"" (Van Loon et al., 2014, p. 2139). They see PUOH gesture as a pragmatic marker of turn taking which has a potential to be grammaticalized to act as a grammatical marker of cohesion (e.g., signaling parallel clausal structures to ensure connection between the ideas).

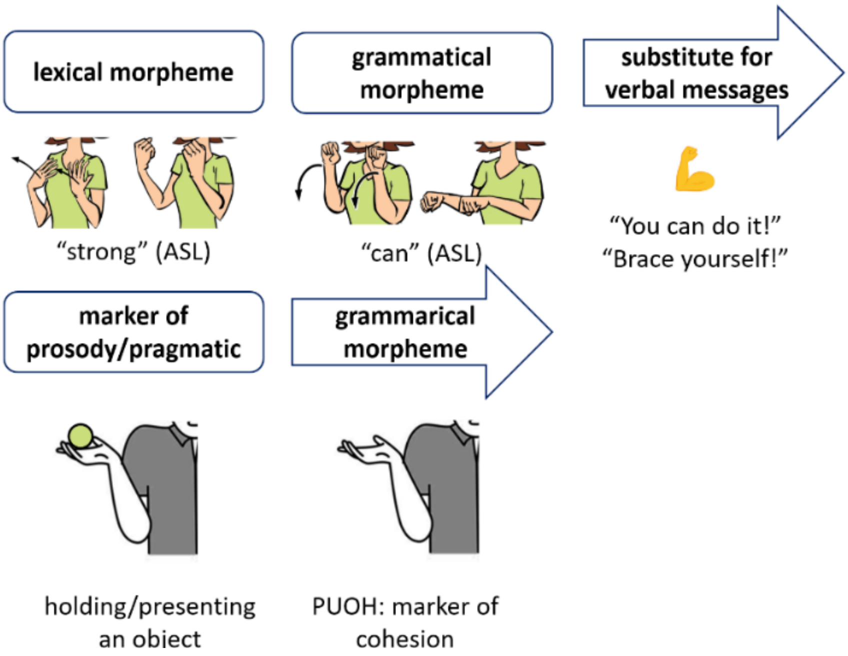


Figure 5.10.Two routes of gesture grammaticalization (adapted from Ladewig, 2020; Van Loon et al., 2014)

Other examples of gestures entering the grammatical system of a language are cases of negation and plurality already mentioned in the previous section. For instance, all of the gestures in the family of Away gestures are not only bound together by the similarity in form (rapid decisive movement away from the body as if clearing the space), but also the commonality of themes the represent, namely, "rejection, refusal, negative assessment, and negation" (Bressem & Müller, 2014b, p. 1596). While a lot of researchers focused on the integration of these gestures in the structure of verbal utterance aiming to see how gesture phases and elements in the verb phrase are coordinated as, for instance, presented in the article by Harrison (2010) on palm down horizontal across body gesture (PDacross). However, we also had an example of BAS gesture in gesture sequences used as a meaning-bearing element in the absence of speech.

Another notion, plurality, has been briefly outlined in this paper as one of the instances of reduplication. From the perspective of grammaticalization, the potential of gestures to convey plurality lies in the principle diagrammatic iconicity and reflects the pattern "more form = more meaning" (Bressem, 2020, p. 1). While gestures as a marker of plurality seem to always occur together in speech and, thus, can be treated as redundant, the consistency of the interplay between gesture and speech in expressing plurality suggests the existence of multimodal. Moreover, while conveying similar meaning, both gesture and speech operate in different dimensions, which makes "a verbo-gestural conceptualization of plurality (…) multiplex, discrete, and bounded" ((Talmy, 2000, as cited in Bressem, 2020, p. 15). Again, similar to the notion of aspect, the way plurality is construed in speech-accompanying gesture follows the similar structural principle as observed in sign languages, such as ASL.

The topic of grammaticalization of recurrent gestures that have been extensively researched in recent years. Three grammatical functions taken over by gestures are frequently referred to: negation, aspect, and plurality. These examples show that although these grammatical functions have always been seen as the phenomena of a spoken language, they should rather be seen as modality-independent due to the potential of gestures (both in sign languages and as bodily movements coupled with speech) to convey them independently of speech.

## 5.3 Conclusion: dependence or autonomy

The chapter focused on outlining the syntactic properties of gesture. Its goal was to show different approaches to the analysis of these properties as compared to language in order to see whether the system of co-speech gesture solely feeds on the system of spoken language, or it is capable of generating an autonomous system of its own.

On the one hand, the patterns of gesture integration show that gestures are highly dependent on the structures generated by spoken language. It also concerns the case of the independent use of co-speech gestures in the absence of speech. Such examples are rare and are likely to show the potential of spontaneous gesture to transition into symbolic signs. On the other hand, a great overlap in the meaning and structure bearing features in spoken language and co-speech gestures allows for establishing various patterns of their interaction and extend the understanding of gesture beyond single-unit grammars. For instance, there is a set of examples of code manifestation (section 5.2) that shows the potential for compositional structure in co-speech gesture. These examples also show that gestures can sometimes transcend their use as a complementary tool and display features of an independent system.

Nevertheless, it may be useful to stick to the middle ground. Considering the powerful nature of spoken language, it is unreasonable to attempt to assign all its properties to co-speech gesture; rather, it is necessary to distinguish which of these properties belong to the general principles of cognition spanning different types of human behaviors, which reside in perception, which ones constitute gesture proper, and which are the result of gesture-speech interface occurring on-line in the process of meaning-making.

# 6 Multimodal Parallel Architechture

The Parallel Architecture (henceforth PA) is an approach to linguistic theory which evolved from the new understanding of phonology and semantics developed in the 1970s. It was proposed by American linguist Ray Jackendoff, who, among others, sought to give an answer to the questions posed by Chomsky and other generativists earlier on regarding the physiological and biological nature of the faculty of language. However, the machinery suggested by Jackendoff went against the mainstream generative approach to language. Instead of placing emphasis on syntax as a main source of combinatorial structure of language from which both phonology and semantics are derived, Jackendoff argued that different structures, namely, phonology, syntax, and semantics are "independent generative components (…) each with its own primitives and principles of combination" (Jackendoff, 2010, p. 1). These independent structures are coordinated by interface rules, principles that establish optimal linkages between structures of two types (Figure 6.1). In other words, these rules are not derivational, they are rather treated as (potentially violable) constraints, which posit no inherent directionality or logical sequence on the construction of complex representations.

According to Jackendoff (2004), the core of the process of human conceptualization lies in the principles of lexical decomposition, which, in turn, represents a rich system of differentiation and interplay between the notions of Conceptual/Semantic Structure (CS) and Spatial Structure (SpS), focal values, cluster concepts, semantic fields, function-argument structure, qualia structure, and dot-object structure. In this section, the notions of CS and SpS, semantic fields, and function-argument structure will be discussed together with their application to the semantics of gesture.

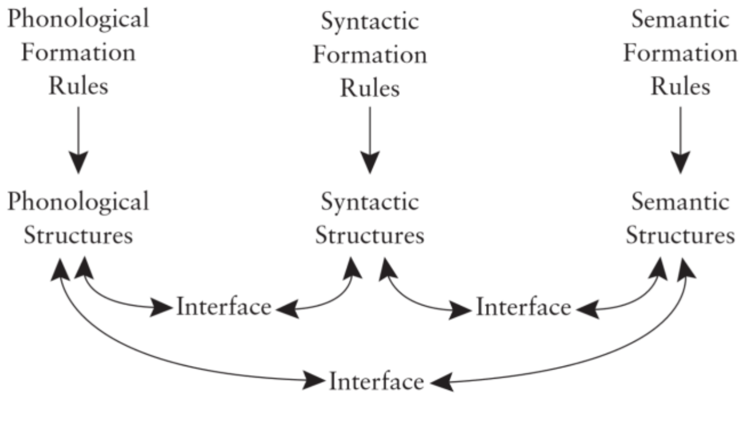


Figure 6.1. Jackendoff's Parallel Architecture (Jackendoff, 2010, p. 1)

Another important notion, developed by Jackendoff is the idea of Unification (as opposed to Merge) as a basic computational operation, which is responsible for clipping pieces of stored structure in the process of generating novel sequences. Moreover, Jackendoff claims that Unification is "a general brain mechanism for achieving combinatoriality" that "can be generalized to combinatorial cognitive capacities other than language" (Jackendoff, 2010, p. 27). This makes the PA a highly flexible framework, which can also provide an account on how theory of language can be extended to other human capacities. The idea of independent representations, mediated by interface principles, reflects other processes that take place in the human mind.

Firstly, Jackendoff claims that lexical items as well as lexical concepts are stored in long-term memory. He also argues that the lexical concepts are not monadic and innate (Fodor, 1998, 2010), but must have a compositional structure. Moreover, these primitives the meaning can be decomposed to can be abstract and, hence, are not solely based on perceptual mechanisms and sensory information, as "perceptual descriptive features alone are not sufficient for characterizing meaning; and there is no way to construct inferential descriptive features from perceptual primitives" (Jackendoff, 2004, p. 339). This leads to a division of the structure of meaning into two major components: Conceptual Structure (CS) and Spatial Structure (SpS).

CS is understood as a "hierarchical arrangement built out of discrete features and functions; it encodes such aspects of understanding as category membership (taxonomy) and predicate-argument structure". CS forms a basis for reference. It deals with formally linguistic notions of predicate-argument relations, semantic categories, type-token distinction, etc. The above-outlined idea of what CS comprises also goes in line with Gärdenfors's theory of conceptual spaces (Gärdenfors, 2004) which unites other theories such as Lakoff and Johnson's "image schemas" (Lakoff & Johnson, 2003), Zlatev's "mimetic schemas" (Zlatev, 2014; Zlatev et al., 2005), as well as notions of metaphor and metonymy. Moreover, it has been shown that gestures similar to spoken language are grounded in the same principles of organization of conceptual representations, which suggests that cognitive mechanisms underlying human communicative capacity are modality independent.

In turn, SpS is related to the system of vision, which also "receives and integrates inputs about shape and spatial layout from the haptic system (sense of touch), auditory localization, and the somatosensory system" (Jackendoff, 2004, p. 346). SpS supports object identification and categorization, gives information regarding possible variations of the object shape, and encodes the overall spatial layout. In other words, it can be treated as a part of the system of general cognition. Here, Jackendoff draws a clear line between the notions of image and SpS: image is restricted to a given instance rooted in a visual modality, while SpS uses information obtained via proprioception, not directly accessible by vision. This idea originally draws on Marr's idea of 3D model (Marr, 2010). However, in contrast to Marr's model, SpS codes not only static but also dynamic configurations. As has been seen in the previous chapters, our embodied experiences with the world play a significant role in how we structure conceptualization and encode our thoughts in both language and gesture. Here, the affordances of bodily modality due to its ability to simultaneously encode different types of information from various dimensions as well as its iconic transparency allow for establishing direct link between the meaning and form of a gesture and perceptual input potentially bypassing the stage of conceptualization (the CS). There are also components that allow for interfacing between CS and SpS, such as "notions of physical object, part-whole relationships, locations, force, and causation" (Jackendoff, 2004, p. 347) all of them present in both structures. Crucially for the co-speech gesture, as has been shown in chapter 3 (Semantics), SpS is also responsible for encoding image-schemas that represent one of the forces that motivate the choice of gestural form. For instance, talking about semantic fields, Jackendoff voices an important idea, that should be taken into account when discussing the potential of gesture integration into the analysis of language. He says that "many grammatical patterns used to describe physical objects in space also appear in expressions that describe non-spatial domains" (Jackendoff, 2004, p. 356). Lakoff and Johnson (Lakoff, 1994; Lakoff & Johnson, 2003) see such phenomenon as a part of a system of metaphor, interpreting metaphor as any type of extension from one domain to another. Jackendoff's perspective is different, as he contends that such cases reflect the processes of the basic functioning of human cognition, which "permits complex thought to be formulated and basic entailments to be derived in any domain" (Jackendoff, 2004, p. 358). The spatial domain is considered to be frequently used in a vast majority of languages due to its evolutionary priority. Thus, his argument is that although metaphor is ubiquitous in thought, such phenomena should be seen as "a set of precise abstract underlying conceptual patterns that can be applied to many different semantic fields" (Jackendoff, 2004, p. 358).

As has been shown, following Jackendoff's Parallel Architecture, it is possible to accommodate the domain of gesture into the model and establish connections between new components to show how grammar and meaning interact across different modalities.

The idea was initially suggested by Cohn and Schilperoord, who, in line with Jackendoff's theory, defined three mutually interfacing structural components of multimodal architecture: modality, meaning and grammar (Cohn, 2016; Cohn & Schilperoord, 2022). Modality is seen as a "channel by which a message is expressed and conveyed" (Cohn & Schilperoord, 2022, p. 3). It involves a "cluster of substructures that include the sensory apparatus for producing and comprehending a signal, the cognitive structures that guide such signals, and the combinatorial principles that govern how those signals combine" (Cohn & Schilperoord, 2022, p. 3). Gestures represent bodily modality, produced through articulations of the different parts of the body, which makes it possible to look at bodily movements as an equivalent of phonological structure. They are connected to syntactic and conceptual structures via interface links, as is shown in Figure 6.2:

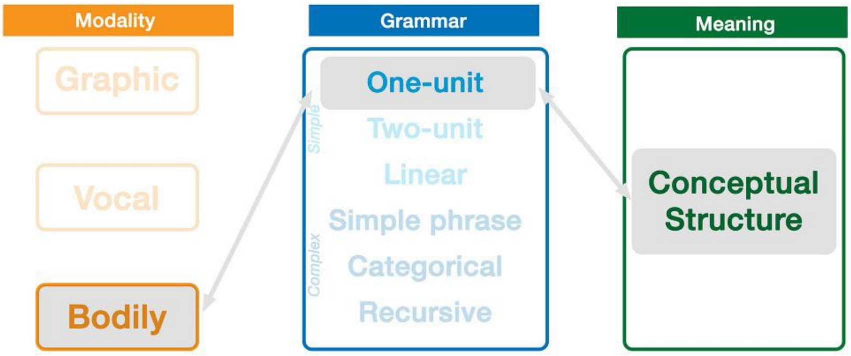


Figure 6.2. Activation of unimodal expression: bodily modality (Cohn & Schilperoord, 2022, p. 4)

According to Figure 6.2, even within one modality expressions can be encoded into three fundamental components of the PA. It means that since such interactions between these three components are available to all the modalities, correspondences between different modalities are also possible. These interactions occur on the level of conceptual structure, while both the modalities and grammars can remain relatively independent. A more complex model for a co-speech gesture including both vocal and bodily modalities and reflecting interactions between them is shown in Figure 6.3.

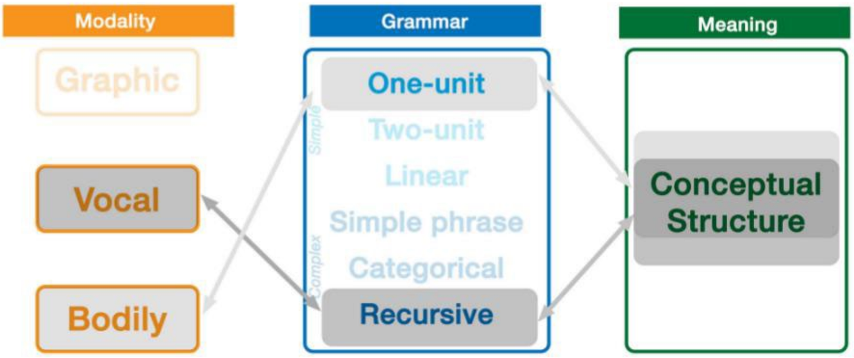


Figure 6.3. Multimodal activation (Cohn & Schilperoord, 2022, p. 4)

The scheme in Figure 6.3. shows how simultaneous activation of two unimodal expressions, such as the interaction of a grammatically complex sentence in natural language together with one-unit gesture, leads to the emergence of a single multimodal expression. The proposed model treats different structures engaged in the course of interaction as the elements of a "single, holistic communicative system for conceptual expression" (Cohn, 2016, p. 308), encompassing different types of information used to decode the meaning and similarity between linguistic/grammatical structures underlying it. Within the framework of multimodal PA gestures may be seen as constituents within complex linguistic and/or gestural phrases. In this case, the interface rule for one-unit grammar in bodily modality can be the following:

Modality: $[_{\text{Utterance}}$ Gesture unit$_1]_2$
Syntax/grammar: $[\text{GU}_1]_2$
Semantics/meaning: $[\text{F}\ (\text{X}_1)]_2$

As Cohn and Schilperoord (C&S henceforth) put it, ""language" is not an amodal representation that "flows out" of different modalities, but rather all modalities are present and persisting as part of a larger holistic communicative faculty" (Cohn & Schilperoord, 2022, p. 5). This raises a natural question on how the interplay between different modalities and their grammar happens to be subsequently turned into a single unified multimodal unit.

In this thesis, based on the evidence and the analysis provided in the chapters above, we propose an extended model of multimodal PA (Figure 6.4) focusing on vocal and gestural modalities. Although the model is based on the model proposed by C&S (2022), it differs in the following aspects. Firstly, while C&S only assign syntactic organization of gestures to one-unit grammars, we argue that gestural organization can potentially go beyond one-unit grammars. As has been shown in chapter 5 (Syntax), gestures can form linear sequences and have a potential to be interpreted as recursive. Importantly, as has been shown in the example of gestural reduplications and gestural sequences, gestures have the capacity to obey the principle of compositionality. Moreover, as illustrated in chapter 3 (Morphology) on the examples of phonesthemes which are argued to function as elements of recurrent gestures and gesture families. Gestures are similar to language even on lower levels of language and display some features of rudimentary morphology. At the same time, we admit that interpretation of syntactic features of gestures is coupled with much ambiguity: although gestures share many principles of syntactic organization with grammars of signed and spoken languages, they utilize different modality-specific resources (such as space) to build a structure. Moreover, they are unlikely to be as rich and independent. Secondly, we have developed the idea of what elements shared between spoken language and gesture may constitute Conceptual Structure (CS). This list is by no means meant to be exhaustive: nevertheless, such elements of conceptual structure as image schemas, mimetic schemas, metaphor and metonymy stem, according to the argument developed in this work, from modality-conditioned perceptual symbol systems, which is supported by the body of evidence provided in Chapter 2 (Spatial Structure and Conceptual Structure) and Chapter 4 (Semantics). It does not, however, mean that these concepts are manifested equally in both modalities. This idea is similar to the notion of logical form versus grammatical encoding where logical form remains the same across contexts and specifies the same global meaning, while grammatical encoding brings different aspect to the foreground with respect to the system and affordances of a given language. That is, although speech and gesture share the same underlying cognitive processes that participate in meaning conceptualization, they highlight different features of these mental representations. Moreover, C&S left out Spatial Structure from their discussion with the purpose of doing more profound research before drawing any conclusions. In the current work we abstain from theorizing much further and restrict SpS to the domain of vision and spatial cognition as one the most important sources guiding our experiences and providing rich information about the surrounding physical world. Nevertheless, establishing an interface link between the mental representation of language and perception is crucial to address the question how we can communicate about what we experience.  Lastly, the model in Figure 6.4. reflects one of the most important features of communication, namely, the fact that language is variably multimodal. This means that the structures within the model are not linearly connected and, thus, their simultaneous presence at all times is not obligatory for both modalities.

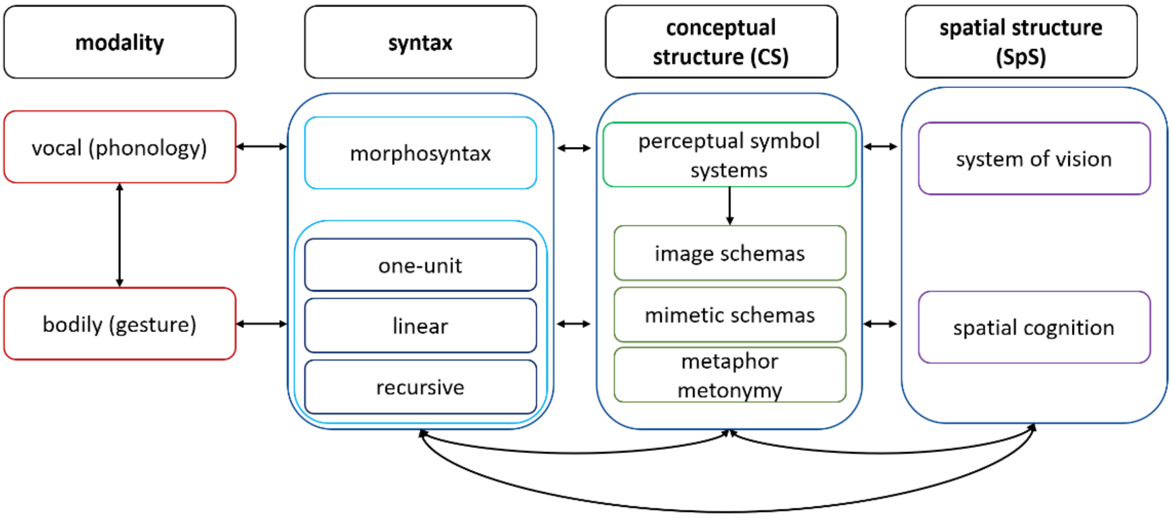


Figure 6.4. Multimodal Parallel Architecture (co-speech gesture oriented)

The grammars of spoken languages are very prominent. As a result, the research on gesture which started much later traditionally focused on defining similarities and differences within a spoken language paradigm. The current thesis provides an overview of the existing approaches to research on speech-entangled gestures which shows that so far researchers have taken the knowledge about spoken languages as a baseline to talk about other phenomena occurring in communicative contexts. This does not necessarily mean that such approach is wrong, however it may be useful to have a different perspective on the matter. The framework of multimodal Parallel Architecture (PA) proposed in this chapter can serve as one of the main theoretical models for the new research. Importantly, the model does not have explanatory power: at the current stage of multimodal research, it cannot provide any definite answer neither regarding the extent of the linguistic potential of gesture, nor the evolution of language. Rather, the model represents a highly flexible system that can account for the variety of resources deployed in the process of communication. Moreover, the multimodal PA leaves enough room for interpretation; it can both support the view of the system of gesture where semantic and syntactic structures follow their own generative principles, and the idea that the system of gesture is not elaborate enough to generate any structures of their own and can only be integrated into the system established in a spoken language. Alternatively, within the model, it can be argued that in some instances gestures can bypass the system of spoken language, by deriving their form-meaning relationships from the autonomous system of perceptual symbols. In other words, the PA can accommodate both scenarios of co-speech gesture as a simple but at the same time autonomous structure and a system which feeds off the resources determined and refined by the highly elaborate system of spoken language. However, it is important to avoid the trap of going into the detail of what these scenarios constitute, because it will lead to unavoidably pouring from empty into void due to the lack of any definite evidence for either. One of the greatest problems is that, as has been shown, language and gesture share the same (or very similar) principles of building Conceptual structure (CS) and Spatial structure (SpS). This makes it extremely difficult to draw conclusions of what is present in spoken language that was influenced by the extended use of gestures and vice versa.

# 7 Discussion and conclusion

One may think that due to the extensive reliance on spoken language as a main tool for communication that is invariably present in humans' daily interaction (among members of the society who communicate using verbal modality), other semiotic resources employed in the course of interaction cannot be as prominent and, thus, are secondary in the linguistic research: meaning can be very efficiently conveyed without access to, for instance, visual modality. However, it is undeniable, that language is inherently multimodal and that modalities are tightly linked to one another while at the same time displaying certain unique traits. In the current thesis, the main focus was on bodily modality and, in particular, on gestures used together with speech. The main difficulty while discussing the role of these gestures is to describe their complex and versatile nature, but at the same time not to overstate their potential.

Borrowing the term introduced within Construction Grammar, it is possible to imagine the existence of multimodal constructions, stable form-meaning pairings where form can be shared between several modalities. How do they function? Language *is* multimodal by nature; however, it is also important that it is *variably* multimodal. This means that the use of co-speech gestures or, more broadly, multimodal constructions, is not obligatory. But "obligatoriness should not be used as the only criterion for identifying multimodal constructions" (Hoffmann, 2021, p. 85). From what we know from the existing research on perception and cognition, it is possible to say that we have permanent access to sensory information about facts and objects of the physical world based on our direct and/or indirect experience with those. And, according to the theory of Perceptual Symbol Systems (PSS) this information is modality dependent. So, the question posed in this work is not whether or not multimodal information is present in mental representations stored in a long-term memory, but rather how this information is used by people when performing both linguistic and non-linguistic task (e.g., mental manipulations). Is it activated simultaneously all the time? Are some responses more automatic than others (e.g., common vs uncommon instances in language)? Are there such things as multimodal constructions stored in the long-term memory or is multimodality an artefact of cognitive operations in the working memory occurring online?

In order to answer these questions, further research should focus more on the nature of multimodality as a cognitive phenomenon linked to both perception and language rather than as an observable behavior. It also may be useful to look at co-speech gestures not as integral components of language, but as a system in its own right that can, together with spoken language, serve as a communicative tool. Thus, the goal of the emerging theories should be to unite the existing views on verbal and gestural production in order to create a framework that would explain how composite utterances function within a shared language ecology.

In this thesis, it was proposed that one of the existing frameworks, namely, the adaptation of Jackendoff's Parallel Architecture (PA), is flexible enough to accommodate the wide array of resources deployed in the process of communication starting with conceptual structure of human minds and finishing with the observable emergent behaviors. By introducing an elaborated framework of multimodal PA, the thesis also attempted to answer the question regarding the similarities between gestures and spoken language, also using findings from sign language research. From the PA perspective it seems that the structural overlap may exist due to two factors. First, both verbal and bodily modality have common conceptual system guided by the properties of general cognition. In this case it is possible to argue that co-speech gestures have a potential to bypass the spoken language system. In other words, language records our experiences with the world in its semantic and syntactic structure, and so do gestures, but they do it differently. That is, the language system has analogues for the perceived objects, events, and states as well as the ability to reason, process of introspection, etc. both on the level of semantics and grammar/syntax. As a result, the production of verbal and gestural signs is largely guided by the same cognitive principles. This can cause redundancy; however, different modalities tend to highlight particular aspects of objects/events/experiences/etc., depending on their affordances (e.g., different ways of talking about spatial events employed by language and gesture). Secondly, because of the dominant position of spoken communication, the use of co-speech gestures can resemble the structure of spoken communication since the interfaces between them strengthened over the centuries of co-existence.

The framework of PA can also account for the structural properties of sign languages. It has been shown by a few decades of sign language research that sign language is a fully fledged language; albeit having many structural similarities with spoken language, it constitutes a system of its own (e.g., sign languages do not resemble their counterpart spoken languages). This can be explained by the similar notion of interfaces between autonomous systems residing in perception, spoken modality, and bodily modality. It is possible to argue that the scenario of gestures manifesting a simpler but autonomous structure is more plausible from an evolutionary perspective, since such structures can be found in monkeys who are the great source of evidence for the presence of linguistic capacities which can be extended to early hominins. As a result, the multimodal PA can provide perspective to answering questions about cognitive evolution and the functioning of language in the brain.

Many theories and approaches described in this work (such as the theory of Perceptual Symbol Systems (PSS), Conceptual Spaces, multimodal image schemas, etc.) as well as the framework of multimodal PA proposed by the author of this work require extensive empirical research in order to take off. It becomes clear that in order to investigate the process of "languaging" and attempt to answer the question regarding how general (non-linguistic) abilities such as memory, imagery, and visual attention are related to the use of language, more principled cross-linguistic research is required that will focus on the cross-linguistic comparison between speech, co-speech gestures, and signs in sign language. Secondly, the existing knowledge highlights the need for an experimental paradigm that will investigate the relations between perceptual domains and conceptual structure observed through the use of language and gesture.

To sum up, the framework of multimodal Parallel Architecture offers a compelling explanation of how different modalities function together, which can be supported by indirect evidence as described in the current work. Nowadays, no one doubts that co-speech gestures often become parts of communicative act (speaker's communicative intent) but without a strong empirical support the main questions still remain open: should co-speech gestures be treated as an integral but autonomous part of language as communicative system and, if yes, what makes them so?

# References

Anourova, I., Nikouline, V. V., Ilmoniemi, R. J., Hotta, J., Aronen, H. J., & Carlson, S. (2001). Evidence for Dissociation of Spatial and Nonspatial Auditory Information Processing. *NeuroImage*, *14*(6), 1268–1277. https://doi.org/10.1006/nimg.2001.0903

Arnott, S. R., Binns, M. A., Grady, C. L., & Alain, C. (2004). Assessing the auditory dual-pathway model in humans. *NeuroImage*, *22*(1), 401–408. https://doi.org/10.1016/j.neuroimage.2004.01.014

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, *22*(4), 577–660. https://doi.org/10.1017/S0140525X99002149

Bavelas, J. B., Chovil, N., Lawrie, D. A., & Wade, A. (1992). Interactive gestures. *Discourse Processes*, *15*(4), 469–489. https://doi.org/10.1080/01638539209544823

Bavelas, J., Gerwing, J., Sutton, C., & Prevost, D. (2008). Gesturing on the telephone: Independent effects of dialogue and visibility. *Journal of Memory and Language*, *58*(2), 495–520. https://doi.org/10.1016/j.jml.2007.02.004

Bedny, M., Caramazza, A., Grossman, E., Pascual-Leone, A., & Saxe, R. (2008). Concepts Are More than Percepts: The Case of Action Verbs. *The Journal of Neuroscience*, *28*(44), 11347–11353. https://doi.org/10.1523/JNEUROSCI.3039-08.2008

Bellmund, J. L. S., Gärdenfors, P., Moser, E. I., & Doeller, C. F. (2018). Navigating cognition: Spatial codes for human thinking. *Science*, *362*(6415), eaat6766. https://doi.org/10.1126/science.aat6766

Birdwhistell, R. L. (1971). *Kinesics and Context: Essays on Body Motion Communication*. University of Pennsylvania Press. https://doi.org/10.9783/9780812201284

Bohn, M., Call, J., & Tomasello, M. (2019). Natural reference: A phylo- and ontogenetic perspective on the comprehension of iconic gestures and vocalizations. *Developmental Science*, *22*(2). https://doi.org/10.1111/desc.12757

Boutet, D., Blondel, M., Beaupoil-Hourdel, P., & Morgenstern, A. (2021). A multimodal and kinesiological approach to the development of negation in signing and non-signing children. *Languages and Modalities*, *1*, 31–47. https://doi.org/10.3897/lamo.1.68150

Bressem, J. (2013). 70. A linguistic perspective on the notation of form features in gestures. In C. Müller, A. Cienki, E. Fricke, S. Ladewig, D. McNeill, & S. Tessendorf (Eds.), *Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science (HSK) 38/1* (pp. 1079–1098). DE GRUYTER. https://doi.org/10.1515/9783110261318.1079

Bressem, J. (2014). 124. Repetitions in gesture. In C. Müller, A. Cienki, E. Fricke, S. Ladewig, D. McNeill, & J. Bressem (Eds.), *Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science (HSK) 38/2* (pp. 1641–1649). DE GRUYTER. https://doi.org/10.1515/9783110302028.1641

Bressem, J. (2020). Conceptualizing plurality as bounded areas in space–Reduplication and diagrammatic iconicity as semiotic forces in multimodal language use. *Body Diagrams: On the Epistemic Kinetics of Gesture*.

Bressem, J. (2021). *Repetitions in Gesture: A Cognitive-Linguistic and Usage-Based Perspective*. De Gruyter. https://doi.org/10.1515/9783110697902

Bressem, J., & Müller, C. (2014a). 119. A repertoire of German recurrent gestures with pragmatic functions. In C. Müller, A. Cienki, E. Fricke, S. Ladewig, D. McNeill, & J. Bressem (Eds.), *Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science (HSK) 38/2* (pp. 1575–1591). DE GRUYTER. https://doi.org/10.1515/9783110302028.1575

Bressem, J., & Müller, C. (2014b). 120. The family of Away gestures: Negation, refusal, and negative assessment. In C. Müller, A. Cienki, E. Fricke, S. Ladewig, D.

McNeill, & J. Bressem (Eds.), *Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science (HSK) 38/2* (pp. 1592–1604). DE GRUYTER. https://doi.org/10.1515/9783110302028.1592

Bush, D., Barry, C., Manson, D., & Burgess, N. (2015). Using Grid Cells for Navigation. *Neuron*, *87*(3), 507–520. https://doi.org/10.1016/j.neuron.2015.07.006

Casasanto, D. (2008). *Conceptual affiliates of metaphorical gestures*. International Conference on Language, Communication, & Cognition. Brighton, UK.

Chater, N., & Christiansen, M. H. (2010). Language Acquisition Meets Language Evolution. *Cognitive Science*, *34*(7), 1131–1157. https://doi.org/10.1111/j.1551-6709.2009.01049.x

Chevrefils, L., Danet, C., Doan, P., Thomas, C., Rébulard, M., Contesse, A., Dauphin, J.-F., & Bianchini, C. S. (2021). The body between meaning and form: Kinesiological analysis and typographical representation of movement in Sign Languages. *Languages and Modalities*, *1*, 49–63. https://doi.org/10.3897/lamo.1.68149

Christiansen, M. H., & Chater, N. (2022). *The language game: How improvisation created language and changed the world*. Basic Books.

Cienki, A. (2005a). Image schemas and gesture. *From Perception to Meaning: Image Schemas in Cognitive Linguistics*, *29*, 421–442.

Cienki, A. (2005b). Metaphor in the '"Strict Father"' and '"Nurturant Parent"' cognitive models: Theoretical issues raised in an empirical study. *Cogl*, *16*(2), 279–312. https://doi.org/10.1515/cogl.2005.16.2.279

Cienki, A. (2009). Conceptual Metaphor Theory in Light of Research on Speakers' Gestures. *Cognitive Semiotics*, *5*(1–2), 349–366. https://doi.org/10.1515/cogsem.2013.5.12.349

Cienki, A. (2013). 11. Cognitive Linguistics: Spoken language and gesture as expressions of conceptualization. In C. Müller, A. Cienki, E. Fricke, S. Ladewig, D. McNeill, & S. Tessendorf (Eds.), *Handbücher zur Sprach- und Kommunikationswissenschaft /*

*Handbooks of Linguistics and Communication Science (HSK) 38/1* (pp. 182–201). DE GRUYTER. https://doi.org/10.1515/9783110261318.182

Cienki, A. (2017). *Ten lectures on spoken language and gesture from the perspective of cognitive linguistics: Issues of dynamicity and multimodality*. Brill.

Cienki, A. (2022). The study of gesture in cognitive linguistics: How it could inform and inspire other research in cognitive science. *WIREs Cognitive Science*, *13*(6). https://doi.org/10.1002/wcs.1623

Cienki, A. (2023). *Recurrent gestures with pragmatic functions*. VU WInter School.

Cienki, A., & Müller, C. (Eds.). (2008). *Metaphor and Gesture* (Vol. 3). John Benjamins Publishing Company. https://doi.org/10.1075/gs.3

Clark, H. H. (1996). *Using Language* (1st ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511620539

Cloutman, L. L. (2013). Interaction between dorsal and ventral processing streams: Where, when and how? *Brain and Language*, *127*(2), 251–263. https://doi.org/10.1016/j.bandl.2012.08.003

Cochet, H., & Vauclair, J. (2010). Pointing gesture in young children: Hand preference and language development. *Gesture*, *10*(2–3), 129–149. https://doi.org/10.1075/gest.10.2-3.02coc

Cohn, N. (2016). A multimodal parallel architecture: A cognitive framework for multimodal interactions. *Cognition*, *146*, 304–323. https://doi.org/10.1016/j.cognition.2015.10.007

Cohn, N., & Schilperoord, J. (2022). Remarks on Multimodality: Grammatical Interactions in the Parallel Architecture. *Frontiers in Artificial Intelligence*, *4*, 778060. https://doi.org/10.3389/frai.2021.778060

Corballis, M. C. (2017). A Word in the Hand: The Gestural Origins of Language. In M. Mody (Ed.), *Neural Mechanisms of Language* (pp. 199–218). Springer US. https://doi.org/10.1007/978-1-4939-7325-5_10

Coulson, S. (2001). *Semantic Leaps: Frame-Shifting and Conceptual Blending in Meaning Construction* (1st ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511551352

Dehaene, S., Meyniel, F., Wacongne, C., Wang, L., & Pallier, C. (2015). The Neural Representation of Sequences: From Transition Probabilities to Algebraic Patterns and Linguistic Trees. *Neuron*, *88*(1), 2–19. https://doi.org/10.1016/j.neuron.2015.09.019

Desimone, R. (1990). Neural mechanisms of visual processing in monkeys. *Handbook of Neuropsychology*.

Emmorey, K., Winsler, K., Midgley, K. J., Grainger, J., & Holcomb, P. J. (2020). Neurophysiological Correlates of Frequency, Concreteness, and Iconicity in American Sign Language. *Neurobiology of Language*, *1*(2), 249–267. https://doi.org/10.1162/nol_a_00012

Enfield, N. J. (2013). 44. A "Composite Utterances" approach to meaning 45. Towards a grammar of gestures: A form-based view. In C. Müller, A. Cienki, E. Fricke, S. Ladewig, D. McNeill, & S. Tessendorf (Eds.), *Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science (HSK) 38/1* (pp. 689–707). DE GRUYTER. https://doi.org/10.1515/9783110261318.689

Fasolo, M., & D'Odorico, L. (2012). Gesture-plus-word combinations, transitional forms, and language development. *Gesture*, *12*(1), 1–15. https://doi.org/10.1075/gest.12.1.01fas

Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex (New York, NY: 1991)*, *1*(1), 1–47.

Fibigerová, K., & Guidetti, M. (2022). Iconicity in gesture: How Czech children and adults use iconic gestures to deal with a gap between mental and linguistic representations of motion events. In S. Lenninger, O. Fischer, C. Ljungberg, & E. Tabakowska (Eds.), *Iconicity in Language and Literature* (Vol. 18, pp. 245–264). John Benjamins Publishing Company. https://doi.org/10.1075/ill.18.12fib

Fitch, W. T. (2014). Toward a computational framework for cognitive biology: Unifying approaches from cognitive neuroscience and comparative cognition. *Physics of Life Reviews*, *11*(3), 329–364. https://doi.org/10.1016/j.plrev.2014.04.005

Fodor, J. A. (1998). *Concepts: Where Cognitive Science Went Wrong* (1st ed.). Oxford University PressOxford. https://doi.org/10.1093/0198236360.001.0001

Fodor, J. A. (2010). *The language of thought* (Digit. repr). Harvard Univ. Pr.

Fricke, E. (2013). 46. Towards a unified grammar of gesture and speech: A multimodal approach. In C. Müller, A. Cienki, E. Fricke, S. Ladewig, D. McNeill, & S. Tessendorf (Eds.), *Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science (HSK) 38/1* (pp. 733–754). DE GRUYTER. https://doi.org/10.1515/9783110261318.733

Fricke, E. (2014a). 122. Kinesthemes: Morphological complexity in co-speech gestures. In C. Müller, A. Cienki, E. Fricke, S. Ladewig, D. McNeill, & J. Bressem (Eds.), *Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science (HSK) 38/2* (pp. 1618–1630). DE GRUYTER. https://doi.org/10.1515/9783110302028.1618

Fricke, E. (2014b). 125. Syntactic complexity in co-speech gestures: Constituency and recursion. In C. Müller, A. Cienki, E. Fricke, S. Ladewig, D. McNeill, & J. Bressem (Eds.), *Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science (HSK) 38/2* (pp. 1650–1661). DE GRUYTER. https://doi.org/10.1515/9783110302028.1650

Fricke, E., Bressem, J., & Müller, C. (2014). 123. Gesture families and gestural fields. In C. Müller, A. Cienki, E. Fricke, S. Ladewig, D. McNeill, & J. Bressem (Eds.), *Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science (HSK) 38/2* (pp. 1630–1640). DE GRUYTER. https://doi.org/10.1515/9783110302028.1630

Fröhlich, M., Wittig, R. M., & Pika, S. (2019). The ontogeny of intentional communication in chimpanzees in the wild. *Developmental Science*, *22*(1), e12716. https://doi.org/10.1111/desc.12716

Gärdenfors, P. (2004). *Conceptual spaces: The geometry of thought*. MIT Press.

Gärdenfors, P. (2014). *The geometry of meaning: Semantics based on conceptual spaces*. MIT press.

Gärdenfors, Peter (Director). (2022, November 20). *Conceptual Spaces and the Geometry of Word Meanings*. https://www.youtube.com/watch?v=87O9fnu8BTU

Gibbs, Jr., R. W. (Ed.). (2008). *The Cambridge Handbook of Metaphor and Thought* (1st ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511816802

Grady, J. E. (1997). *Foundations of meaning: Primary metaphors and primary scenes*. University of California, Berkeley.

Graham, K. E., Hobaiter, C., Ounsley, J., Furuichi, T., & Byrne, R. W. (2018). Bonobo and chimpanzee gestures overlap extensively in meaning. *PLOS Biology*, *16*(2), e2004825. https://doi.org/10.1371/journal.pbio.2004825

Hafri, A., Green, E. J., & Firestone, C. (2023). *Compositionality in visual perception* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/trg7q

Hagoort, P., & Van Berkum, J. (2007). Beyond the sentence given. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1481), 801–811. https://doi.org/10.1098/rstb.2007.2089

Harrison, S., & Ladewig, S. H. (2021). Recurrent gestures throughout bodies, languages, and cultural practices. *Gesture*, *20*(2), 153–179. https://doi.org/10.1075/gest.21014.har

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The Faculty of Language: What Is It, Who Has It, and How Did It Evolve? *Science*, *298*(5598), 1569–1579. https://doi.org/10.1126/science.298.5598.1569

Hodge, G., & Ferrara, L. (2022). Iconicity as Multimodal, Polysemiotic, and Plurifunctional. *Frontiers in Psychology*, *13*, 808896. https://doi.org/10.3389/fpsyg.2022.808896

Hoffmann, T. (2021). Multimodal Construction Grammar. In *The Routledge Handbook of Cognitive Linguistics*. Routledge.

Iriskhanova, O. K., & Cienki, A. (2018). THE SEMIOTICS OF GESTURES IN COGNITIVE LINGUISTICS: CONTRIBUTION AND CHALLENGES. *Voprosy Kognitivnoy Lingvistiki*, *4*, 25–36. https://doi.org/10.20916/1812-3228-2018-4-25-36

Iverson, J. M., & Goldin-Meadow, S. (2005). Gesture Paves the Way for Language Development. *Psychological Science*, *16*(5), 367–371. https://doi.org/10.1111/j.0956-7976.2005.01542.x

Jackendoff, R. (1985). *Semantics and cognition* (Nachdr.). MIT Pr.

Jackendoff, R. (2004). *Foundations of language: Brain, meaning, grammar, evolution* (1. publ. in paperback, 2. impr). Oxford Univ. Press.

Jackendoff, R. (2010). *Meaning and the lexicon: The parallel architecture, 1975-2010*. Oxford University Press.

Jakobson, R., & Halle, M. (1956). Fundamentals of language.'s-Gravenhage: Mouton. *The Hague: Mouton & Co*.

Johnson, M. (1987). *The Body in the Mind: The Bodily Basis of Meaning, Imagination, and Reason*.

Jones, S. S., & Smith, L. B. (1993). The place of perception in children's concepts. *Cognitive Development*, *8*(2), 113–139. https://doi.org/10.1016/0885-2014(93)90008-S

Kaye, L. K., Malone, S. A., & Wall, H. J. (2017). Emojis: Insights, affordances, and possibilities for psychological science. *Trends in Cognitive Sciences*, *21*(2), 66–68.

Kendon, A. (1995). Gestures as illocutionary and discourse structure markers in Southern Italian conversation. *Journal of Pragmatics*, *23*(3), 247–279. https://doi.org/10.1016/0378-2166(94)00037-F

Kendon, A. (2004). *Gesture Visible Action as Utterance* (1st ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511807572

Kendon, A. (1988). *How gestures can become like words.* This paper is a revision of a paper presented to the American Anthropological Association, Chicago, Dec 1983.

Kendon, A., R. Key, M., & M. Harris, R. (Eds.). (1975). *Organization of behavior in face-to-face interaction*. Mouton [u.a.].

Kersken, V., Gómez, J.-C., Liszkowski, U., Soldati, A., & Hobaiter, C. (2019). A gestural repertoire of 1-to 2-year-old human children: In search of the ape gestures. *Animal Cognition*, *22*, 577–595.

Kita, S. (2000). How representational gestures help speaking. In D. McNeill (Ed.), *Language and Gesture* (pp. 162–185). Cambridge University Press. https://doi.org/10.1017/CBO9780511620850.011

Kita, S., & Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal?: Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and Language*, *48*(1), 16–32. https://doi.org/10.1016/S0749-596X(02)00505-3

Kockelman, P. (2005). The semiotic stance. *Semiotica*, *2005*(157), 233–304. https://doi.org/10.1515/semi.2005.2005.157.1-4.233

Kok, K. (2016). The grammatical potential of co-speech gesture: A Functional Discourse Grammar perspective. *Functions of Language*, *23*(2), 149–178. https://doi.org/10.1075/fol.23.2.01kok

Kubicek, E., & Quandt, L. C. (2021). A Positive Relationship Between Sign Language Comprehension and Mental Rotation Abilities. *The Journal of Deaf Studies and Deaf Education*, *26*(1), 1–12. https://doi.org/10.1093/deafed/enaa030

Ladewig, S. H. (2011). Putting the cyclic gesture on a cognitive basis. *CogniTextes*, *6*(Volume 6). https://doi.org/10.4000/cognitextes.406

Ladewig, S. H. (2014a). 118. Recurrent gestures. In C. Müller, A. Cienki, E. Fricke, S. Ladewig, D. McNeill, & J. Bressem (Eds.), *Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science (HSK) 38/2* (pp. 1558–1574). DE GRUYTER. https://doi.org/10.1515/9783110302028.1558

Ladewig, S. H. (2014b). 121. The cyclic gesture. In C. Müller, A. Cienki, E. Fricke, S. Ladewig, D. McNeill, & J. Bressem (Eds.), *Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication*

*Science (HSK) 38/2* (pp. 1605–1618). DE GRUYTER. https://doi.org/10.1515/9783110302028.1605

Ladewig, S. H. (2014c). 126. Creating multimodal utterances: The linear integration of gestures into speech. In C. Müller, A. Cienki, E. Fricke, S. Ladewig, D. McNeill, & J. Bressem (Eds.), *Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science (HSK) 38/2* (pp. 1662–1677). DE GRUYTER. https://doi.org/10.1515/9783110302028.1662

Ladewig, S. H. (2020). *Integrating gestures: The dimension of multimodality in cognitive grammar*. De Gruyter Mouton.

Lakoff, G. (1987). Cognitive models and prototype theory. *Cambridge University Press*, 63–100.

Lakoff, G. (1994). What is a conceptual system. *The Nature and Ontogenesis of Meaning*, 41–90.

Lakoff, G., & Johnson, M. (2003). *Metaphors we live by*. University of Chicago Press.

Langacker, R. (2008). *Cognitive Grammar: A Basic Introduction* (1st ed.). Oxford University PressNew York. https://doi.org/10.1093/acprof:oso/9780195331967.001.0001

Langacker, R. W. (1986). An Introduction to Cognitive Grammar. *Cognitive Science*, *10*(1), 1–40. https://doi.org/10.1207/s15516709cog1001_1

Langacker, R. W. (2014). *Foundations of cognitive grammar. Vol. 1: Theoretical prerequisites* (Nachdr., Vol. 1). Stanford Univ. Press.

Leyton, M. (2004). *A Generative Theory of Shape* (Vol. 2145). Springer Berlin Heidelberg. https://doi.org/10.1007/3-540-45488-8

Liebal, K. (2014). *Primate communication: A multimodel approach*. Cambridge University Press.

MacNeilage, P. F. (2010). *The origin of speech* (Issue 10). Oxford University Press.

Marr, D. (2010). *Vision: A computational investigation into the human representation and processing of visual information*. MIT Press.

McNeill, D. (1985). So you think gestures are nonverbal? *Psychological Review*, *92*(3), 350–371. https://doi.org/10.1037/0033-295X.92.3.350

McNeill, D. (1992). Hand and Mind: What Gestures Reveal about Thought. *Leonardo*, *27*(4), 358. https://doi.org/10.2307/1576015

McNeill, D. (2005). *Gesture and Thought*. University of Chicago Press. https://doi.org/10.7208/chicago/9780226514642.001.0001

McNeill, D., & Duncan, S. D. (2000). Growth points in thinking-for-speaking. *Language and Gesture*, *1987*, 141–161.

McNeill, D., & Levy, E. (1982). Conceptual representations in language activity and gesture. *Speech, Place, and Action*, 271–295.

Mishra, R. K., & Marmolejo-Ramos, F. (2010). On the mental representations originating during the interaction between language and vision. *Cognitive Processing*, *11*, 295–305.

Mittelberg, I. (2008). Peircean semiotics meets conceptual metaphor: Iconic modes in gestural representations of grammar. In A. Cienki & C. Müller (Eds.), *Gesture Studies* (Vol. 3, pp. 115–154). John Benjamins Publishing Company. https://doi.org/10.1075/gs.3.08mit

Mittelberg, I. (2013). 47. The exbodied mind: Cognitive-semiotic principles as motivating forces in gesture. In C. Müller, A. Cienki, E. Fricke, S. Ladewig, D. McNeill, & S. Tessendorf (Eds.), *Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science (HSK) 38/1* (pp. 755–784). DE GRUYTER. https://doi.org/10.1515/9783110261318.755

Mittelberg, I. (2014). 130. Gestures and iconicity. In C. Müller, A. Cienki, E. Fricke, S. Ladewig, D. McNeill, & J. Bressem (Eds.), *Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science (HSK) 38/2* (pp. 1712–1732). DE GRUYTER. https://doi.org/10.1515/9783110302028.1712

Mittelberg, I., & Waugh, L. R. (2009). Chapter 14. Metonymy first, metaphor second: A cognitivesemiotic approach to multimodal figures of thought in co-speech gesture.

In C. J. Forceville & E. Urios-Aparisi (Eds.), *Multimodal Metaphor* (pp. 329–358). Mouton de Gruyter. https://doi.org/10.1515/9783110215366.5.329

Müller, C. (1998). *Redebegleitende Gesten: Kulturgeschichte, Theorie, Sprachvergleich*. Spitz.

Müller, C. (2004). Forms and uses of the Palm Up Open Hand. A case of a gesture family? *Semantics and Pragmatics of Everyday Gestures*, 234–256.

Müller, C. (2014). 128. Gestural modes of representation as techniques of depiction. In C. Müller, A. Cienki, E. Fricke, S. Ladewig, D. McNeill, & J. Bressem (Eds.), *Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science (HSK) 38/2* (pp. 1687–1702). DE GRUYTER. https://doi.org/10.1515/9783110302028.1687

Müller, C. (2018). Gesture and Sign: Cataclysmic Break or Dynamic Relations? *Frontiers in Psychology*, *9*, 1651. https://doi.org/10.3389/fpsyg.2018.01651

Müller, C., Bressem, J., & Ladewig, S. H. (2013). 45. Towards a grammar of gestures: A form-based view. In C. Müller, A. Cienki, E. Fricke, S. Ladewig, D. McNeill, & S. Tessendorf (Eds.), *Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science (HSK) 38/1* (pp. 707–733). DE GRUYTER. https://doi.org/10.1515/9783110261318.707

Müller, C., Ladewig, S. H., & Bressem, J. (2013). 3. Gestures and speech from a linguistic perspective: A new field and its history. In C. Müller, A. Cienki, E. Fricke, S. Ladewig, D. McNeill, & S. Tessendorf (Eds.), *Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science (HSK) 38/1* (pp. 55–81). DE GRUYTER. https://doi.org/10.1515/9783110261318.55

Novack, M. A., & Goldin-Meadow, S. (2017). Gesture as representational action: A paper about function. *Psychonomic Bulletin & Review*, *24*(3), 652–665. https://doi.org/10.3758/s13423-016-1145-z

Özçalişkan, Ş., Gentner, D., & Goldin-Meadow, S. (2014). Do iconic gestures pave the way for children's early verbs? *Applied Psycholinguistics*, *35*(6), 1143–1162. https://doi.org/10.1017/S0142716412000720

Özyürek, A., Willems, R. M., Kita, S., & Hagoort, P. (2007). On-line Integration of Semantic Information from Speech and Gesture: Insights from Event-related Brain Potentials. *Journal of Cognitive Neuroscience*, *19*(4), 605–616. https://doi.org/10.1162/jocn.2007.19.4.605

Parker, G. J. M., Luzzi, S., Alexander, D. C., Wheeler-Kingshott, C. A. M., Ciccarelli, O., & Lambon Ralph, M. A. (2005). Lateralization of ventral and dorsal auditory-language pathways in the human brain. *NeuroImage*, *24*(3), 656–666. https://doi.org/10.1016/j.neuroimage.2004.08.047

Peeters, D., Snijders, T. M., Hagoort, P., & Özyürek, A. (2017). Linking language to the visual world: Neural correlates of comprehending verbal reference to objects through pointing and visual cues. *Neuropsychologia*, *95*, 21–29. https://doi.org/10.1016/j.neuropsychologia.2016.12.004

Peirce, C. S. (1955). *Philosophical writings of Peirce* (Vol. 217). Courier Corporation.

Pfau, R., & Steinbach, M. (2006). *Modality-independent and modality-specific aspects of grammaticalization in sign languages*. Univ.-Verl.

Piaget, J. (1972). Intellectual Evolution from Adolescence to Adulthood. *Human Development*, *15*(1), 1–12. https://doi.org/10.1159/000271225

Pike, K. L. (1967). *Language in Relation to a Unified Theory of the Structure of Human Behavior:* DE GRUYTER. https://doi.org/10.1515/9783111657158

Planer, R. J., & Sterelny, K. (2021). *From Signal to Symbol: The Evolution of Language*. The MIT Press. https://doi.org/10.7551/mitpress/13906.001.0001

Pouw, W., de Wit, J., Bögels, S., Rasenberg, M., Milivojevic, B., & Ozyurek, A. (2021). Semantically Related Gestures Move Alike: Towards a Distributional Semantics of Gesture Kinematics. In V. G. Duffy (Ed.), *Digital Human Modeling and Applications in Health, Safety, Ergonomics and Risk Management. Human Body, Motion and*

*Behavior* (Vol. 12777, pp. 269–287). Springer International Publishing. https://doi.org/10.1007/978-3-030-77817-0_20

Pouw, W., Dingemanse, M., Motamedi, Y., & Özyürek, A. (2021). A Systematic Investigation of Gesture Kinematics in Evolving Manual Languages in the Lab. *Cognitive Science*, *45*(7). https://doi.org/10.1111/cogs.13014

Pouw, W., & Fuchs, S. (2022). Origins of vocal-entangled gesture. *Neuroscience & Biobehavioral Reviews*, *141*, 104836. https://doi.org/10.1016/j.neubiorev.2022.104836

Rauscher, F. H., Krauss, R. M., & Chen, Y. (1996). Gesture, speech, and lexical access: The role of lexical movements in speech production. *Psychological Science*, *7*(4), 226–231.

Ruth-Hirrel, L., & Wilcox, S. (2018). Speech-gesture constructions in cognitive grammar: The case of beats and points. *Cognitive Linguistics*, *29*(3), 453–493. https://doi.org/10.1515/cog-2017-0116

Sablé-Meyer, M., Ellis, K., Tenenbaum, J., & Dehaene, S. (2021). *A language of thought for the mental representation of geometric shapes* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/28mg4

Saussure, F. de. (2005). *Course in general linguistics* (C. Bally, Ed.; Nachdr.). McGraw-Hill.

Schlenker, P., & Chemla, E. (2018). Gestural agreement. *Natural Language & Linguistic Theory*, *36*(2), 587–625. https://doi.org/10.1007/s11049-017-9378-8

Schwartz, R. (1995). Is Mathematical Competence Innate? *Philosophy of Science*, *62*(2), 227–240. https://doi.org/10.1086/289854

Secora, K., & Emmorey, K. (2020). Visual-Spatial Perspective-Taking in Spatial Scenes and in American Sign Language. *The Journal of Deaf Studies and Deaf Education*, *25*(4), 447–456. https://doi.org/10.1093/deafed/enaa006

Senghas, A., & Coppola, M. (2001). Children Creating Language: How Nicaraguan Sign Language Acquired a Spatial Grammar. *Psychological Science*, *12*(4), 323–328. https://doi.org/10.1111/1467-9280.00359

Seyfeddinipur, M. (2004). *Meta-discursive gestures from Iran: Some uses of the 'Pistolhand.'*

Stephens, D., & Tuite, K. (1983). *The hermeneutics of gesture*. Paper presented at the symposium on Gesture at the meeting of the american Anthropological Association, Chicago.

Stokoe, W. C. (1960). *Sign language structure: An outline of the visual communication systems of the American deaf*. 1–78.

Streeck, J. (2009). *Gesturecraft: The manu-facture of meaning* (Vol. 2). John Benjamins Publishing Company. https://doi.org/10.1075/gs.2

Streeck, J. (2013). 43. Praxeology of gesture. In C. Müller, A. Cienki, E. Fricke, S. Ladewig, D. McNeill, & S. Tessendorf (Eds.), *Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science (HSK) 38/1* (pp. 674–688). DE GRUYTER. https://doi.org/10.1515/9783110261318.674

Sweetser, E. (1998). *Regular metaphoricity in gesture: Bodily-based models of speech interaction*. Actes du 16e Congrès International des Linguistes.

Taglialatela, J. P., Russell, J. L., Schaeffer, J. A., & Hopkins, W. D. (2011). Chimpanzee Vocal Signaling Points to a Multimodal Origin of Human Language. *PLoS ONE*, *6*(4), e18852. https://doi.org/10.1371/journal.pone.0018852

Talmy, L. (1988). Force Dynamics in Language and Cognition. *Cognitive Science*, *12*(1), 49–100. https://doi.org/10.1207/s15516709cog1201_2

Talmy, L. (2000). *Toward a cognitive semantics. Cambridge (Mass.)*.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of Visual and Linguistic Information in Spoken Language Comprehension. *Science*, *268*(5217), 1632–1634. https://doi.org/10.1126/science.7777863

Taub, S. F. (2001). *Language from the Body: Iconicity and Metaphor in American Sign Language* (1st ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511509629

Teßendorf, S. (2014). 117. Pragmatic and metaphoric – combining functional with cognitive approaches in the analysis of the "brushing aside gesture." In C. Müller, A. Cienki, E. Fricke, S. Ladewig, D. McNeill, & J. Bressem (Eds.), *Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science (HSK) 38/2* (pp. 1540–1558). DE GRUYTER. https://doi.org/10.1515/9783110302028.1540

Tomasello, M. (2010). *Origins of human communication* (1. MIT paperback ed). MIT Press.

Van Loon, E., Pfau, R., & Steinbach, M. (2014). 169. The grammaticalization of gestures in sign languages. In C. Müller, A. Cienki, E. Fricke, S. Ladewig, D. McNeill, & J. Bressem (Eds.), *Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science (HSK) 38/2* (pp. 2133–2149). DE GRUYTER. https://doi.org/10.1515/9783110302028.2133

Volterra, V., Roccaforte, M., Di Renzo, A., & Fontana, S. (2022). *Italian Sign Language from a Cognitive and Socio-semiotic Perspective: Implications for a general language theory* (Vol. 9). John Benjamins Publishing Company. https://doi.org/10.1075/gs.9

Wilcox, S. (2017). ROUTES FROM GESTURE TO LANGUAGE. *Revista Da ABRALIN*, *4*(1/2). https://doi.org/10.5380/rabl.v4i1/2.52651

Willems, K., & De Cuypere, L. (Eds.). (2008). *Naturalness and Iconicity in Language* (Vol. 7). John Benjamins Publishing Company. https://doi.org/10.1075/ill.7

Winter, B. (2021). Iconicity, not arbitrariness, is a design feature of language. *Talk Presented for the Abralin Talk Series, May*, *5*.

Wolf, D., Rekittke, L.-M., Mittelberg, I., Klasen, M., & Mathiak, K. (2017). Perceived Conventionality in Co-speech Gestures Involves the Fronto-Temporal Language Network. *Frontiers in Human Neuroscience*, *11*, 573. https://doi.org/10.3389/fnhum.2017.00573

Wundt, W. (1900). *Völkerpsychologie. Eine Untersuchung der Entwicklungsgesetze von Sprache, Mythos und Sitte*. Engelmann, Leipzig.

Zelinsky-Wibbelt, C. (1983). *Die semantische Belastung von submorphematischen Einheiten im Englischen: Eine empirisch-strukturelle Untersuchung*. P. Lang.

Zlatev, J. (2007). Intersubjectivity, Mimetic Schemas and the Emergence of Language. *Intellectica. Revue de l’Association pour la Recherche Cognitive*, *46*(2), 123–151. https://doi.org/10.3406/intel.2007.1281

Zlatev, J. (2014). Image schemas, mimetic schemas and children’s gestures. *Cognitive Semiotics*, *7*(1), 3–29. https://doi.org/10.1515/cogsem-2014-0002

Zlatev, J., Persson, T., & Grdenfors, P. (2005). Bodily mimesis as the missing link in human cognitive evolution, series no. 121. *Lund University Cognitive Studies.[JZ]*.

NTNU

Norwegian University of
Science and Technology