Human Multimodal Communication

*From Concrete Processing Systems to the Abstract Organisation of Language*

B.A. Essay

Klara Sif Skarphéðinsdóttir

Kt.: 191289-2509

Supervisors: Matthew James Whelpton
Jesper Greve Kristensen

May 2017
Abstract

The origins of the human language capacity is a much debated topic among scholars. In the late nineteenth century the structure of human thought was believed to be the key to discovering the structure of language and culture. Today, however, research shows the opposite to be true. The neurological foundations for social cognition are the true catalysts for the subsequent linguistic abilities of humans. In this thesis I show that language builds upon imitative foundations that underlie socialisation and culture. These foundations rely on intentional, coordinated movements of the hand to the mouth, present from the fetal stage in both human and non-human primates, and on the ability to cognitively mirror such movements in others. This coordination of movements is believed to be the basis for action-planning and goal-directed thought, which support the development of grammar and syntax. Usually, the debate centres around whether speech stems from gestural origins, with movement through a protolinguistic stage towards language, or whether it was a sudden change in human cognition which rendered evolutionary studies of communication obsolete. However, gesture itself is multifaceted and can be either pre-linguistic or co-speech. Thus, even gesture, which is physical and a concrete mode of communication, can take symbolic forms when coupled with language. In all forms of communication, context is essential for meaning-making. Meaning arises before an understanding of conceptual signs, and has its foundations in cognitive imitation, or mirroring. The imitative approach to the study of language reunites the dualisation created in Western thought of mind from body. It makes the previously mind-independent sign once again dependent on the body. At all levels of culturalisation there is a movement from lower-level concrete processing systems towards higher-level, abstract organisational systems. Thus, the body is the key to language, and the abstract is brought back into the concrete.
1. Introduction..............................................................................................................1
2. Evolutionary Origins................................................................................................4
   2.1 Human Protolanguage................................................................................4
   2.2 Evidence from Non-Human Primates........................................................5
   2.3 Intention and Mirroring..............................................................................6
3. Communicative Foundations....................................................................................8
   3.1 Low-Level Imitation...................................................................................8
   3.2 Attention.....................................................................................................9
   3.3 Triadic Mimesis........................................................................................11
4. Linguistic Processing..............................................................................................12
   4.1 Conceptual Grounding..............................................................................12
   4.2 Abstract Knowledge.................................................................................14
   4.3 Episodic and Semantic Knowledge..........................................................15
5. Understanding Speech............................................................................................17
   5.1 Higher-Level Imitation.............................................................................17
   5.2 Discourse Coherence................................................................................18
   5.3 Co-Speech Gestures..................................................................................20
6. Conclusion..............................................................................................................22
Reference List.............................................................................................................24
1. Introduction

“...language provides an immensely delicate and useful way of pointing.”
(p. 234, de Bruin & de Haan, 2013).

A human's ability to utilise a complex, symbolic system of communication is what sets them apart from other primates. Mimesis, also known as 'mental mimicry' or simply 'imitation', is the facilitator of the emergence of language, due to its role in the initial structuring of human culture and the subsequent structure of grammar and syntax. The study of language, its nature, structure, and supporting cognitive faculty, has been the focus of inquiry within linguistics for many decades. Ultimately, the question of language concerns the concept of the sign, and how words and mental representations came to be 'about' anything. There are many theories of sign production and interpretation within humans and other non-human primates. In the late 1800s, Ferdinand de Saussure described the study of signs as a branch of psychology, naming the field semiology (Barbieri, 2009). His account of the nature of a sign characterised it as a dual entity: the signifier and the signified, or the code and the message. Saussure also made a distinction between la langue (language) and parole (speech). According to Saussure, la langue is the most important of all sign systems due to its semantic universality, or its ability to communicate information about all possible scenarios, people, places, or things, imaginary or real (Chandler, 2000). Saussure's dichotomy of the signifier and signified, and of la langue and parole, caused the conception of the sign to become abstracted from physical experience and described it as independent from the concrete. This follows from contemporary ideas of the role of binary oppositions in the structure of human thought and culture, resulting in the separation of the mind from the body within Western thought systems. In this way, Saussure reduced the linguistic system to a fixed, synchronic entity that is understood outside of any context. The belief was that the sign's meaning could only be found in relation to other signs – one sign is only understood as a specific concept because it is not something else, and these meanings were thought to be stable and fixed (Keane, 2003; Threadgold, 2013).

Saussure was criticised for prioritising structure over usage, not taking into account the fact that usage determines structure: the social context affects communicative meaning (Chandler, 2000).

Noam Chomsky had similar ideas as Saussure, in that concepts, according to
Chomsky, are meaningless symbols. However, Chomsky believed that those symbols are not explicitly signs, for signs carry meaning. To him, syntax and structure comes before conceptual meaning because words are meaningless abstract symbols, sentences are strings of such symbols, and language itself is simply a set of such strings (Lakoff, 2012). Chomsky believed that the origin of language was irrelevant. He believed that the specificity of language, and syntax in particular, in humans meant that there is an inherent divide between us and other animals, and as such, evolution over time plays no part. Furthermore, as language and its structure evolved only recently in human history, Chomsky proposed that the capacity for syntax and grammar appeared due to a genetic mutation in the human genome, and thus are not applicable to a Darwinian theory of language evolution. Many scholars call this change 'the great leap forward' (Barbieri, 2010; Corballis, 2014). The arbitrary nature of words and concepts – the fact that they have no resemblance to the outside, external reality – was also one of the reasons that Chomsky believed in the 'great leap forward'. Environmental factors were thought to have no part in shaping the structure of language. To Chomsky, thought, or internal language, evolved before external, or communicative language. Structure is therefore understood before meaning. In this way, external language depends on a human's ability to understand what others are thinking, or their internal language. Chomsky developed a theory of 'Universal Grammar', which he believed to be similar to a specific organ that developed due to the linguistic mutation. Within this line of thought, universal grammar is what allows humans to understand language: the sentence's literal interpretation being computed through this capacity by combining the fixed word meanings, information from previous conversations, and general world knowledge in order to formulate a coherent and shared meaning (Corballis, 2014; Cosentino, 2014). According to Charles Darwin, in contrast to Chomsky's position, the differences between humans and other animals are not of type, but of degree, and language is not an exception. The language faculty in humans can be seen as a specific trait, and not a separative characteristic (Ferretti & Adornetti, 2014). As opposed to Chomsky's idea of a primary internal language existing before a secondary, external language, the fact is that language is always learned in a face-to-face, external context, due to external communicative pressures.

Besides Saussure's ideas about the duality of signs, another school of semiotics existed, led by Charles Sanders Peirce. His idea of the sign held that there was not a duality lying with the signifier and the signified, but a triad that lies within the sign as a
*representamen*, or the form that the sign takes whether material or not, an *interpretant*, or the way the sign is seen or made sense of, and the *object*, or that to which the sign refers. The interaction between the three is called by Peirce 'semiosis'. When comparing Saussure's signifier and signified with Peirce's model, a similarity can be seen between the representamen and signifier, and the interpretant and signified. The interpretant, however, is different from the signified in that it is a sign in and of itself (Chandler, 2000). “The sign does not give direct access to objects. It mediates that relation by giving rise to further signs.” (p. 413, Keane, 2003). In this way, 'semiosis' is a *process* that takes place first and foremost among and in relation to other human beings.

Corballis (2014) believes that the structure of our thoughts is what gave rise to the structured nature of language, and that this structure was achieved through increasingly complex organizational skills that developed first and foremost through the making of complex tools. The skills needed for such activities require an ability to imitate others at a higher level than other non-human primates are able to. They require mimesis, which is fundamentally a socially acquired ability. Mimesis, or the ability to imitate others mentally and physically, develops through the relationship between infant and caretaker and later expands through relationships with others. In contrast to the idea that structure arises before meaning, meaning is primary and gives life to the sign. In this thesis, I will argue in favour of the view that language is initially built upon mimesis, and that knowing the meaning of a sign does not require language or syntax, but an understanding of the sign's presence along with the intention to communicate something. In this way, both gestural and linguistic communication are built on lower-level, concrete processes that facilitate processing within a higher-level, abstract organizational system.

Section 2 begins with an outline of the communicative structures of non-human primates. Because humans and non-human primates have a shared ancestor, a look into the communication systems of apes sheds light on the basic functions of modern human communication systems. Section 3 delves deeper into the foundations of the human's communicative ability, and explains how the first instances of the 'sign' develops in small children. Section 4 describes the neurological basis of language, from the acquisition of concrete concepts to the way abstraction takes place. The section ends with an account of a human's long-term memory and general world knowledge, without which humans could never establish communicative understanding. Finally in Section 5 I show how communicative meaning develops between a speaker and a listener, and
how it is based on attention and the visual availability of co-speech gesture.

2. Evolutionary Origins
2.1 Human Protolanguage
There are many opinions and theories that attempt to explain the origins of human language and culture. The synthetic view of language evolution is one such view, in which single proto-words or gestures are thought to have carried communicative meaning, eventually becoming synthesised into complex, structured utterances. Here, similarly to Chomsky's ideas, the structure of language is understood before meaning can be interpreted. In this view, language understanding takes place mainly at the sentence level based on syntactic structure. Recently, another, holistic view of human language development is favoured, based on gestural and social origins hypotheses (Byrne & Cochet, 2017; Smith, 2006; Corballis, 2014; Cosentino, 2014). In this view, gestures that carried representational meaning, or were iconic, were firstly presented individually as whole complex propositions, lacking in internal morphological structure, such as in pantomime. An analytic stage then follows, in which such propositions were broken down into smaller parts, eventually gaining a systematic structure that became language. Here, meaning is conveyed and understood before structure. This process as a whole is termed 'protolanguage' (Smith, 2006). Human protolanguage was multimodal and used manual, facial and vocal expressions and gestures equally to convey any specific meaning (Arbib, 2013; Quaeghebeur, 2013). Smith (2006) points out that it is possible to combine the two theories to bridge the gap between non-symbolic, iconic representations and modern human language. In this view, more concrete, multimodal ways of communication gave way to a more centralized and abstract mode of communication – from the holistic to the synthetic.

With time and sign-arbitrariness, the vocal modality took on more of the communicative weight than the manual or facial modality. An arbitrary sign refers to any sign that does not relate directly to the meaning that it expresses. For example, in American Sign Language (ASL) the word for 'home' was a combination of hand signs meaning 'sleep' and 'eat'. Today, however, the sign has become more arbitrary, in that the sign for 'eat' is now used near the place where the sign for 'sleep' was, instead of using both signs together. The sign itself, therefore, has become arbitrary, but normative in that its use is understood by the community despite its arbitrariness (Corballis, 2014).
An arbitrary, normative sign is said to be 'symbolic', while pantomimes and less arbitrary gestures are said to be 'iconic'. Arbib (2013) suggests that before sign-arbitrariness, communication consisted of multimodal pantomimes that used the whole body and were indicative of that which they were meant to signify. As the holistic view of language evolution dictates, such signs grew arbitrary and may have developed into a 'protosign'. This is similar to the development of written language – from pictorial hieroglyphs that depicted whole propositions to the analytic stage of breaking up a larger proposition into smaller, more arbitrary ones. In protolanguage, certain vocalisations may at first have been dependent on the gesture that was included, but with time, the gesture may have lost its importance and the word gains independent, arbitrary meaning apart from the gesture it began with, showing therefore a movement from the concrete to the abstract (Gentilucci, Gianelli, Campione & Ferri, 2013). In the deaf and mute, the opposite can be seen: the vocal modality loses any weight it might have had while the gestures became more dominant, systematic, and arbitrary.

2.2 Evidence from Non-Human Primates
In an effort to understand the origins of language and culture, scholars look at non-human primate societies and their modes of communication. The synthetic theory of language evolution, in which it is held that language developed through increasingly complicated vocalisations and a construction of syntax, does not follow the Darwinian theory of evolution. Ape vocalisation is non-volitional and therefore not cultural but innate. Baboons do have hierarchically sanctioned barks and calls, and their vocalisations do vary depending on who is near them, but they are never intentional and sound only during a burst of emotion, such as in fear, anger, or joy. Because of this, their vocalisations are called 'honest signals' (Corballis, 2014; Wacewicz & Zywiczynski, 2008), and are meant for an audience. In some non-human primates, calls do have a specific meaning but it is not possible to attribute semanticity to them such as is found in human language however, because their calls only mean that one specific thing. Some calls, for example, are specified to mean 'leopard', and nothing else (Wacewicz & Zywiczynski, 2008). Ape vocalisations are iconic representations that are encapsulatory and non-volitional. They are to be viewed in tandem with manual and facial gestures – as multimodal.

The holistic, multimodal social origins account of modern human symbolic communication is the perspective that most scholars lean towards today. It has recently
been recorded that most great apes have a repertoire of iconic gestures. Iconic gestures are manual, facial and/or vocal gestures that carry meaning, such as in Arbib's (2013) protosign. Out of 70 to 80 gestures that are included in the African Ape repertoire, around 60 are shared by all of them, and this overlap strongly suggests a common origin (Byrne & Cochet, 2017). Most of these gestures are thought to be innate, although it has been shown time and again that many primates are able to learn more complicated gestures through human intervention. Arbib (2013) calls this 'human-supported ritualisation'. Ritualisation in non-human primates is the process in which the individual learns that an action or behaviour has a role in social communication. It is often used to support the claim that iconicity, though not learned without human intervention, is within a non-human primate's grasp, giving evidence for a possible pre-linguistic mode of meaning-making. Iconicity is a cognitive tool that is essential for communication and is that which allows humans and some apes to refer to things outside of their immediate environment. It is thought that this displacement happens due to increasingly complex social systems (Wacewicz & Zywiczynski, 2008). Studies carried out on baboons have shown that these primates have an intricate understanding of their own social hierarchy and kinship. According to Seyfarth, Cheney and Bergman (2005), iconicity, and then language, developed through knowledge of social relations. The increasing need for cooperation may have led to a need to divide labour between groups and therefore a need to transmit cultural information in order to uphold the system. For this to happen, an individual needs to be able to formulate coherent, and organized plans that lead ultimately to a specific goal (Seyfarth, et. al., 2005; Corballis, 2014).

2.3 Intention and Mirroring
The main difference between an ape's communication system and a human's is that an ape is unable to employ their signs in intentional communication, and cannot use them creatively or in novel ways (Zlatev, 2008; Arbib, 2013). This requires a more complex cognitive system that facilitates intentional and coordinated action-planning in cultural matters, and later in language. The foundation for a functional social institution, such as kinship, is the ability to establish parity between individuals. Parity depends on mutual understanding, which stems from a certain synchronicity that must arise between communicators. Intention, here, is the key. It is the intentionality of an act that imbues it with meaning – especially meaning that is mutually understood by several participants. There are accounts of two classes of neurons found in non-human primates which seem
to act as evolutionary building blocks from which language may have emerged. The first is that which controls grasping in the hands and mouth, and the other controls observation and execution of intentional hand or mouth grasping. The second type is of interest, in that it is thought to be the neurological basis for social interaction, which as stated above is thought to be a precursor to protolanguage. This system of neurons has been termed 'mirror neurons', and was first discovered in the premotor cortex of the macaque monkey brain. It is a system of neurons that becomes active when the individual is executing an intentional action, and also when the individual is observing a con-specific, or a member of one's own species, perform an intentional action (Arbib, 2013; Gentilucci, et. al., 2013) – it is the foundation for imitation and subsequent socialisation.

The reason for the prevalence of the hands and mouth in this region is most likely due to the fact that bringing the hand to the mouth is the first intentional and coordinated action a human and non-human primate executes. The act takes place initially in the womb, grounding the infant in the material world (Needham & Libertus, 2010). Trevarthen (2013) calls this a 'feedback of feeling', and Needham and Libertus (2010) call it 'double-touch'; it is the touch of the infant's hand to its mouth while simultaneously feeling its mouth touch its hand. In non-human primates, such coordinated movements of the hand to the mouth can be elicited when stimulus is applied to the premotor cortex, which implies that the mouth and hands are the tools most prevalent in social communication (Corballis, 2014).

However, despite the fact that humans share these characteristics with other primates, the similarities mostly stop there. Mirror systems in apes and monkeys are rudimentary, and unlike humans, not connected to the vocal modality. When apes vocalise, it is because they are reacting, non-volitionally, to an emotion. The vocalisations, much like a human crying out in pain (or pleasure), are non-intentional and thus are not connected to the intentional brain, or the mirror system. It has been shown in neuroimaging studies that training in vocalisation with human-support does activate this system in apes, but not during communication itself, and hearing the vocalisations of a con-specific does not activate the mirror system (Corballis, 2014; Meguerditchian & Vauclair, 2014). Indeed, the larynx is not under the control of non-human primates, while in humans it clearly is. Though the sound made by a vocalising con-specific does not activate the mirror system in an ape, the sound of manual actions do, such as breaking the shell of a nut, hand-clapping, or lip-smacking (Corballis,
Again, the mirror system activates when one is executing an intentional act oneself, as well as when one is observing (or listening to) another execute such acts. The reason for this mirroring is to understand what the other is doing. The mind must place itself in a state of intending to act in order to understand another's intentional act. Thus, mental-mimicry is a baseline structure for empathy – a purely social phenomenon.

3. Communicative Foundations

3.1 Low-Level Imitation

A homologue to the ape mirror system has been discovered in the human brain, though it is far more complex and extensive than the former. Within this human cortical region are thought to lie the most basic foundations on to which language builds. In Broca's area of the human brain, an area that was originally thought only to relate to language structure and syntax, a region (BA 44) has been found to activate specifically during grasping actions and the observation of grasping actions, very similar to the mirror system of non-human primates (Arbib, 2013; Zlatev, 2008). Area BA 44 in humans specializes in motor-perception-speech functions and it is connected to the premotor cortex. All intentional acts, including communication and articulation in both spoken and signed languages, stimulate activation in these areas (Zlatev, 2008; Arbib, 2013; Meguerditchian & Vauclair, 2014). A human's imitative/intentional system has a much greater differentiation between imitation, observation and execution than the mirror neuron system of apes, and takes place over both hemispheres (Goldman, 2012; Caspers, Zilles, Laird & Eickhoff, 2010). This particular region, BA 44 in Broca's area, passively mirrors another's actions without much conscious effort on the part of the observer. Other's facial expressions, for example, activate certain parts of BA 44 and the premotor cortex, and induces a mirror emotion in the observer, which is believed to play a crucial part in pro-social behaviour (Heyes, 2011; Goldman, 2012). This is thought to be the underlying system that facilitated the evolution of human language. The imitation itself functions in a way to allow for synchronicity between con-specifcics – synchronicity is essential in the building of a social structure, and eventually, human culture (de Bruin & de Haan, 2013).

This level of imitation is termed 'proto-mimesis', and is the type of innate, low-level imitation found in non-human primates and is considered to be parallel to the tongue and lip protrusion that is seen in human infants (Zlatev, 2008; Ray & Heyes,
Proto-mimesis is part of a mimesis hierarchy developed by Jordan Zlatev (2013) that attempts to explain the evolution and development of language in humans. Proto-mimesis consists of a neurological, multimodal mapping of stimuli, in this case vision and kinesthesia, such as the mirror system facilitates. This is sometimes called 'affect perception, or as Zlatev calls it, “empathetic perception,”’ (p. 53). This type of perception is non-intentional and includes infant imitation and emotional reflection in the observer (or listener), and shared attention. Iconic gestures are representational and non-arbitrary. The first iconic gestures that children perform are are said to emerge as “mimetic schemas” (Zlatev, 2014b). Mimetic schemas are pre-verbal foundations for experience-building outside of abstract thought. They are built upon the low-level, body-based imitation as mentioned above. Because a child is surrounded by most of the same people daily, the mimetic schemas are always built upon face-to-face interactions with those people, and are therefore culturally shared. Zlatev (2014b) showed that the iconic gestures of children occur less with speech than other, more abstract gestures such as those that do occur with speech, thus confirming that an individual's first coherent experience in the physical world is built upon a system separate from that of language.

3.2 Attention

The mutual gaze between humans is considered one of the most important underpinnings of communication. Taking place from the moment of birth, shared eye-gaze with the care-givers establishes the basis for joint attention, which is imperative for more complex social interactions (Zlatev, 2008; Felletti, 2014). It is believed that the human eye evolved for the purpose of shared attention, in that the iris and pupil are clearly seen against the white of the eye, and makes it possible for con-specifics to follow the eye-gaze of another despite the position of their heads or body. This is thought to allow for shared attention and cooperation/cohesion between several communicative participants (Perniss & Vigliocco, 2014; Wacewicz & Zywiczynski, 2008; Holler, Schubotz, Kelly, et. al., 2014; Holler, Kokal, Toni, et. al., 2015). During proto-mimesis, eye-gaze on the part of the infant is contagious and non-intentional, much like yawning can be. It is 'synchronous', in that there is no intentional meaning or alterations in gaze. This is the first level of mimicry and is reflected in the mirror-neuron system of non-human primates, as well as in infants through Broca's area BA44. There are two other types of joint attention which develop as the individual matures.
Coordinated joint attention takes place when there are either alterations in gaze or something is being referred to by the child, and develops as the child gains a sense of agency. Reciprocal joint attention is when attention is led and focused by communicative partners. Synchronous joint attention is the only type that has been noted in non-human primates, thus providing evidence of its proto-mimetic status (Zlatev, 2013). During the first months of infancy, mothers spend their time sharing eye-contact with their children, while at the same time imitating the child's facial expressions and engaging in a type of infant-directed speech designed to engage the infant's attention. This supports primary empathy in the form of emotional contagion, in that one's facial expression creates a similar emotion in the observer. This is a part of affect perception, and is a basic, low-level way of creating synchronicity – or shared, mutual understanding – between con-specifics and allows for later, more complex forms of empathy (Zlatev, 2008, 2013; Violi, 2013; Perniss & Vigliocco, 2014).

Mutual gaze leads to a type of attention contagion which is the most simple type of joint attention. This happens when infants from six months of age happen to look in the same direction as the care-giver, non-intentionally – the gaze is contagious. This behaviour has been seen in other non-human primates, and even goats, and is thus part of proto-mimesis (Zlatev, 2008, 2013). Before a human child is able to engage in shared or referential attention, it must be able to differentiate between self and other. This develops through the simplest forms of imitation, attentional and emotional contagion. The child achieves an awareness of referentiality, or meaning, through the care-giver's responses and interpretations of its behaviour. It acts or behaves in a certain way, which the care-giver then reacts to, shifting the “focus from behaviour to what the behaviour means.” (p. 68, Violi, 2013). The child realizes through such interactions that its behaviour means something to the adult, and that it is the focus of attention. Once it understands that its behaviour has an effect on the behaviour of the other, it understands that it is an active agent in the world, and the child gains intentionality and a sense of ownership of its body, separate from that of the care-giver (Zlatev, 2013; Violi, 2013; de Bruin & de Haan, 2013). This leads to a knowledge of goal-oriented behaviour and communicative intent, which are ultimately mediated by mutual gaze. There is a movement away from proto-mimesis into a state of dyadic mimesis, in which conscious control or volition is realized in communicative intent (Zlatev, 2013).
3.3 Triadic Mimesis

When the child understands that it is an active agent, separate from others, the joint attention shared between child and care-giver moves from a simple state of synchronic attention to a state of coordinated attention, and to a point, reciprocal attention. Coordinated attention presents a certain referential gaze, in which the child's gaze alternates between the carer and a referent (Zlatev, 2008, 2013). In most cases this consists of a pointing finger. The child realizes that the other's finger movements are intentional, but does not realize that the finger itself is pointing to an outside object or person. The child imitates the pointing which the adult in turn interprets as having meaning – though the child does not understand that to point is to refer to something in the world particularly, the adult does, and it is the adult's interpretation of the child's movements that eventually creates a referential understanding (Violi, 2013). At around nine months, the child stops looking at the finger itself, and looks instead to where the finger is pointing. This marks the development of reciprocal attention and the onset of triadic mimesis. The child learns that the finger is a sign for something else, and a three way relationship develops between the gesture, its meaning, and the individual for whom the sign is meant (Violi, 2013; de Bruin & de Haan, 2013). Through the action of mutual gaze with the now semiotic child, communicative intent is established. According to Zlatev (2013), the pointing finger is not interpreted as a true sign however, but a 'performative communicative act' that is understood as an intention. This stage is one of multimodal communication, much like the communication of non-human primates. Without human-ritualisation however, it is unavailable to them. It is during the later half of this period that children begin to use the vocal modality more in order to supplement their meaningful gestures rather than to replace them, much as is thought to be the case in the holistic theory of language evolution (Zlatev, 2009).

When the child can jointly attend an object or referent with an adult, it can begin its journey into the symbolic realm of arbitrary and normative language. Joint attention assists the child in making connections between what it sees and handles, and the vocalisations or gestures that accompany it. In fact, a child learns vocabulary much more readily when its attention is already on the object that is being named, and even more readily if the words are accompanied by gestures (Wacewicz & Zywiczynski, 2008; McRae & Jones, 2013; Perniss & Vigliocco, 2014). The fact is, language is always learned in a face-to-face environment due to its multimodality. If a child is not
imitated, for example, it can not imitate; the more one is imitated, the better one becomes at imitating (Ray & Heyes, 2011). Imitation is that which creates synchronicity between con-specifics, and without it, language cannot be learned. At this stage, the mimetic schema, which maps intentional representational acts, begins to give way to a more abstract and linguistic structure. That is, iconic, proto-mimetic gestures that are mapped by mimetic schemas begin to give way to more symbolic, protolinguistic gestures that are mapped in more abstract ways. The mimetic schema is not linguistic, and those bodily gestures that occur without speech are overt, or external, representations of these roots (Zlatev, 2014b). As the child gains experience with the world, and understands that one act stands for, or is similar, to another, the iconic gesture becomes internalized, and can then be reflected upon, or covertly imitated, before being again represented overtly in the absence of that which is being referred to; it becomes displaced from space and time. World-related, concrete concepts are understood in this way, and because of this, many researchers believe that the human motor system plays a vital role in the processing of conceptual, semantic knowledge.

4. Linguistic Processing
4.1 Conceptual Grounding
It is needless to say that the understanding of words requires an ability to remember what it is that they refer to. The human brain houses several types of memory. Working memory is used to process the world as we live it. For example, while walking down a noisy city boulevard, working memory processes all of the sounds, smells, sights, thoughts and emotions that occur at one moment in time. It discards things that are seen as 'background', like the sound of regular traffic, while things that are seen as important, such as meeting a friend unexpectedly, are moved to higher-level processing systems within a long-term memory base. This memory base includes both episodic memory, which includes established memories of one's own past and inner experiences that are connected to that past, and semantic memory, which includes knowledge about concepts such as their meaning and accompanying linguistic signs. Semantic memory has been defined as not solely conceptual knowledge, but as a basis for general world knowledge. It is category-specific, which means that the way our brain organizes the material world around us is highly prototypical and can range from categories of colour, living or non-living entities, body parts, tools, etc. So, a deficit in a specific area within our brains that
stores semantic knowledge about an object, idea, place or person, will result in a
difficulty or inability to distinguish that object, idea, etc., from other, similar objects,
ideas, etc. (Price & Grossman, 2015).

Children in the proto-mimetic stage of linguistic development are innately able
to distinguish between biological motion and non-biological motion. According to Price
and Grossman (2015), this ability is the foundation for semantic categorisation and the
construction of conceptual networks that organize world knowledge into a full cognitive
representation. It has been shown that different aspects of an object activate the
 corresponding areas in the brain, including the visual and motor areas, implying that
semantic/conceptual knowledge is neurologically distributed (Goldman, 2012; Caspers,
et. al., 2010; Price & Grossman, 2015). According to McRae and Jones (2013), when
humans “access word meaning, they automatically activate sensorimotor information
used to perceive and act on the real-world objects and relations to which a word refers.”
(p. 3). There has been a focus on finding quantitative evidence for an essential role of
the motor system in linguistic and conceptual processing. There does indeed seem to be
a certain 'grounding' of concepts that occurs on this level, which can be seen in the
facilitation of children's vocabulary through the help of gesture and pantomime. Thus,
language does seem to build onto a foundation that arises from sensorimotor areas
through mimesis.

There are many differing views on how exactly the brain processes semantic
knowledge. One such view is modality-preferential, in which it is thought that semantic
processing is confined to different spheres, specifically the perceptual and the motor.
Another, similar view, is modality-independent, in which the brain processes semantic
knowledge at a regional level, integrating information from many areas of the brain.
Similarly to Smith's (2006) convergence of the holistic and synthetic views of language
evolution, evidence points to a movement from the modality-preferential to the
modality-independent processing of semantic information (Sakreida, Scorolli, Menz, et.
al., 2013; Handjaras, Ricciardi, Leo, et. al., 2016), or from processing of concrete
concepts in the motor system, to a processing of abstract concepts in semantic memory.
There is much evidence that points to semantic processing in the motor system,
specifically through semantic priming. In a semantic priming experiment, a participant
is exposed to a certain word that is either directed at a body part or an action that uses
the body, such as 'kiss', or 'kick' (Grisoni, Dreyer & Pulvermuller, 2016). The priming
words do not linger long enough for the subject to consciously become aware of them,
but studies show that the brain does perceive such signals. Following the prime is a command for either a congruent or non-congruent action on the part of the participant, which the participant must fulfil immediately. In such experiments it has been shown that non-congruent priming disrupts the movements or utterances of the participant, such as an interrupted grasping motion or a wider mouth in the utterance of a syllable. Conversely, congruent priming facilitates movement – the prime activates the human mirror system and prepares the motor system for movement (Grisoni, et. al., 2016; Heyes, 2011; Gentilucci, et. al., 2013). However, most of these experiments are confined to concrete nouns or verbs: concepts that involve the body or world-experience. Concrete concepts are important however, in that they are mapped directly on to mimetic-schematic representations, resulting in their intimacy with the motor system.

4.2 Abstract Knowledge

Mimetic schemas are body-based representations that take place before, and outside of, abstract thought. When children enter triadic mimesis, they begin to create connections between the world and themselves, not only through their bodies like in mimesis, but also through the aforementioned internalization and displacement of conceptual representations that are connected to those mimetic representations. These connections are said to arise as 'image' or 'perceptual' schemas (Hedblom, Kutz & Neuhaus, 2015; Zlatev, 2014b). These schemas arise from grounded, physical experience within space, in relation to the individual's agency and conceptual knowledge. Through experience with the world, a network, or assemblage of schemas arises that allows for an organisation, recognition, and understanding of a particular scene, word, narrative, or person (Arbib, 2013; Hedblom, et. al., 2015; Preston & Eichenbaum, 2013). In this way, language arises in the individual from an understanding and knowledge of the world and of one's body in that world. Image schemas are the connection between our concrete, sensorimotor experience, and our abstract conceptual knowledge: they make abstraction possible (Arbib, 2013; Hedblom, et. al., 2015). Our abstract conception of 'time', for example, is solely possible due to our concrete knowledge of 'space' – the two are nearly inseparable. According to Irish and Piguet (2013), human semantic memory stems from the fact that the content of experience can be, and is, displaced and made abstract.

There are several theories concerning the precise way in which abstract concepts are mapped. The leading theory, called the dual-coding theory, states that abstract
concepts arise through previous experience with linguistic information and associations, while concrete concepts consist of information and experience from both the sensorimotor system and through linguistic information and associations (Wellsby & Pexman, 2014; Price & Grossman, 2015; McRae & Jones, 2013). In fact, words that are not reflected in the motor system are shown to be specifically categorised in widespread neuronal areas and have therefore probably already been applied to long-term semantic memory. Similarly, Arbib, Gasser & Barrés (2014) suggest that despite the importance of the motor system in the recognition and processing of concrete words, after initial recognition those same motor areas may be 'deemed' unnecessary by the brain in future semantic processing of those specific concepts, and proceed to map directly on to established schematic representations in semantic memory. In this view, language processing is modality-independent. That is, each aspect of an event, situation, person, or object is processed in many different neural modalities, and are then integrated in one particular area to create a whole picture. Therefore, the knowledge that is assembled must exist separately from each physical instantiation in order to be applied to other, related instantiations (Pulvermüller, 2013; Arbib, 2013; Hedblom et al., 2015). The knowledge is therefore abstracted and displaced from the primary content, and abstract concepts can be acquired through direct mapping onto the pre-existing schematic assemblages. This process is extremely important, not only for the learning of concepts, but also to achieve full understanding and communication parity between two or more communicative partners.

4.3 Episodic and Semantic Knowledge
Episodic memory is the memory one has of one's personal experiences. These experiences include the sensory and perceptual aspects and the emotional experience of such aspects. Studies of patients with Alzheimer's Disease have shown that retrieval of episodic memory depends greatly on semantic representations (Irish & Piguet, 2013). According to Irish and Piguet (2013), in order to remember the past or imagine the future, certain contextual factors must be bound or connected to the existing schematic representations that reside within semantic memory. Imagining the future depends on the retrieval of specific details from one's past memories. This is the constructive episodic simulation hypothesis, and it holds that the reason novel situations can be created is because of the creative flexibility of past instantiations (Irish & Piguet, 2013; McRae & Jones 2013). That is, humans are able to retrieve episodic and semantic
details to construct a novel narrative or scenario. This is distinctly connected to the
displacement of concepts from space – a novel experience or object can be reflected on
internally, and then re-represented externally in the absence of said experience or object.
In language, this is expressed as the ability to understand novel sentences not previously
heard before. There is a distinction between episodic memory and semantic memory, in
that episodic memory is one's own lived experience (autobiographical), whereas
semantic memory encompasses cultural and learned experiences. However, this is
precisely what allows humans to acquire, use and understand language so effectively,
especially when the people or sentences that are expressing communicative intent are
novel.

Triadic mimesis allows for an internalization, or a remembrance, of iconic
gestures through the human mirror system. These internalizations can then be retrieved
and reflected on, and subsequently recreated externally with or without the presence of
the original gesture. That is, the representation becomes displaced from space and time
and exists within an image schema. It is a child's care-giver that provides it with the rich
imitative and linguistic material it needs in order to develop into a mature
sociolinguistic being. A human's ability to dissociate and displace situational and
objective elements from the present space and time is what allows the child to
understand that when a non-caregiver directs speech or gesture at it, their bodies and
actions are similar to those of the care-givers. Thus, the child is eventually able to
realize that these imitations and vocalisations are culturally shared. These
developmental epiphanies are an aspect of triadic mimesis – when the children can
interact with more objects or people than solely themselves and those that give them
attention. A child's past knowledge of their care-givers and an ability to mirror their
actions while recognizing a similarity between one and another allows them to
understand the intentions of other people. Ultimately, these experiences with people,
language, and objects become image-schematic representations, which are what allow
for the incorporation of increasingly abstract concepts and the ability to assemble them
in a novel way while still being culturally shared. Thus, the distinction between an
episodic, personal memory-base and culturally-shared semantic knowledge is what
allows for the unification of iconicity and abstraction that ultimately leads to the
symbolic communication system of language.
5. Understanding Speech

5.1 Higher-Level Imitation

The previously mentioned systems of semantic memory are the aspects of human cognition that facilitate higher-level imitation, and therefore establish reciprocal understanding and parity between two communicative partners. As is seen in the social and communicative cognition of monkeys and great apes, their intentional systems are entirely located in their pre-motor cortex. In humans however, the brain's intentional system is a bilateral system that is connected to both Broca's area and the sensorimotor cortex. Unlike Broca's area BA 44, another region within Broca's area (BA 45) controls higher-order linguistic functions (Zlatev, 2008). Non-human primate vocalisation is non-volitional and not affected by their intentional system. BA 45 in humans however, is thought to be specifically related to “articulatory and motor speech functions.” (p. 146, Zlatev, 2008). However, much like area BA 44's connection to the sensorimotor cortex, area BA 45 has also been connected to the premotor cortex, and seems to be the part of the brain that allows humans to imitate and pantomime to the point of creating a complex, symbolic language. According to Zlatev (2013), this type of imitation is essential for the development of human protolanguage, both in the species and in individual development. To a degree, triadic mimesis is available to non-human primates through human ritualization, but the stage of protolanguage is inaccessible to them. In humans, this stage is post-mimetic, in which the vocal modality gains more prominence in communicative acts, and children's vocabulary increases exponentially. As opposed to the lower-level imitation of proto-mimesis and dyadic mimesis, this type of imitation is a higher-level imitation that extends beyond triadic mimesis and is thus part of an increasingly abstract mental representational system.

Many abstract concepts have been shown to be semantically linked to the perception of emotions. As opposed to affect perception, which is an automatic internalization of an external emotional facial or body gesture such as is found in lower-level imitation, emotion perception requires far more complex and abstract knowledge and is intimately tied to higher-level imitation (Pulvermuller, 2013). Attributing complex emotional states to others requires an interaction between long-term semantic and episodic memories, along with the creative ability to combine aspects of each type into a novel idea. It is important to note that ideas are only 'novel' once: schematic assemblages that arise from such an event become applied to long-term semantic
(schematic) memory. Estimating another's intentions is done through both agency attribution, which is a low-level imitative ability acquired during dyadic mimesis, and by using one's own past experience through schematic recollection. When these two aspects are applied together, there is a movement from lower-level, concrete mimicry to higher-level, abstract imitation, also known as interpersonal simulation or theory of mind (Osaka, Ikeda & Osaka, 2012; Shanton & Goldman, 2010).

It is through lower-level affect mimicry that humans are able to build higher-level, abstract conceptualisations about others. Just as abstract concepts are grounded in experience, so are the ways adults interpret others' movements and emotions. Higher-level imitation is far removed from the non-volitional low-level affect mirroring in non-human primates and young children. It depends on the ability to control one's own motor movements, and on being able to plan coherently a sequence of actions that lead towards a certain goal. According to Zlatev (2008), “it is necessary to be able to understand the goal of the modelled action, and to see the action as a means to achieve this goal.” (p. 141). True imitation, like that of a human's capabilities, means being able to imitate to the last detail an action until eventually that action turns into a skill (culturally transmitted) that is transferred to long-term semantic memory. Like when driving a car, much attention to detail is needed during the acquisition of this skill, but eventually the body need not be very closely monitored as one gains experience: the skill is added to our repertoire (Arbib, 2013; Zlatev, 2008). Donald (2012) states that a human's ability to pantomime and gesture is connected to the development of skills because of the same neural mechanisms that are required for both. It is this goal-directedness of our imitative abilities which contributes to discourse coherence, or an understanding of ongoing conversation that includes more than two people.

5.2 Discourse Coherence

If an individual is observing or performing an act that is being carried out with a goal in mind, their brain shifts focus from the visual/perceptual areas towards interpreting an intention or intentions (Osaka, et. al., 2012). Unlike non-human primates, communicative acts activate this intentional brain in humans, which implies certain communicative goals. Since action-planning and execution are connected to the intentional brain, which has its foundations in the planning and execution of reaching and grasping, it is believed that the developing skills required in increasingly complex tool-making are the skills that underlie modern language and discourse coherence.
In patients with traumatic brain injuries, and in some individuals with Autism Spectrum Disorder, executing goal-oriented actions comes with difficulty, and their discourse coherence follows the same path, compromising their ability to stay on topic (Vanegas & Davidson, 2015). Adornetti (2014) believes that there is a moment in pre-historic time in which it is possible to pinpoint the development of the action-planning and goal-oriented behaviour that ultimately facilitated language evolution. Studies in cognitive archaeology have shown that there are neural differences between the production of tools from the Oldowan era versus the production of tools from the later Acheulean era. Archaeologists were asked to produce these tools in the way of early man while under neural observation. Whereas Oldowan tool-making does require motor-manipulation, Acheulean tool-making requires both motor-manipulation and a higher-level of planned, intentional action, which was apparent in the fact that the archaeologists' intentional brain, or mirror system, became active only during the production of the latter, and not the former (Adornetti, 2014; Corballis, 2014).

The foundation of speech understanding depends greatly on the human intentional brain, and the processing of communicative acts relies on volitional control of one's actions and certain organizational functions which have their foundations within semantic memory (Adornetti, 2014). However, before there is an understanding of grammar within a spoken or signed language, there must be an understanding of the meaning that the communicative act is trying to convey. As this is initially made possible through the body, meaning is understood multimodally. This is the level of global coherence, which is an expressive language skill that develops through the observation of facial expressions, posture, arm and hand gesture, and non-linguistic vocalisations (Ellis, Henderson, Wright & Rogalski, 2016). Global coherence does not rely on grammar, due to the fact that coherence at this level occurs pre-linguistically and is thus part of lower-level imitation. So, a structural form of the communicative act is not essential to the coherence of meaning. The act itself, however, and the context it appears in are essential for meaning to be conveyed. In contrast, local coherence is coherence at the level of sentence structure and syntax, and is in line with the synthetic view of language evolution. When language development is observed moving from a global coherence to a local coherence, there becomes a movement from concrete modes of communicative understanding, to more abstract modes of communicative understanding – from the holistic to the synthetic (Adornetti, 2014). Coherence is that
which facilitates the structuring of communicative acts. These acts are then fulfilled by following steps to achieve a goal and ultimately produce an ordered and cohesive narrative or theme.

5.3 Co-Speech Gestures

Language is a multimodal communication system. In signed languages, the manual modality takes precedence over the facial and vocal modalities, and in spoken languages the vocal modality takes over. In both types of communication, the mimetic roots are unmistakeable. Studies have been conducted in which subjects used thought, silent reading, and handwriting – all abstract forms of language – while being recorded through electromyography, using electrodes attached to the skin or muscle to record electrical activity. The results provide the evidence for the mimetic roots of language, in that these different modes of using language all prompted activation in the physiological areas that support speech (Quaeghebeur, 2013). Similarly, when deaf individuals engage in silent reading, the corresponding neural areas that support sign-language become active (Neville, Bavelier, Corina, et. al., 1998). Co-speech gestures in particular have interested researchers in their search for evidence for a gestural foundation of language evolution. These types of gestures are symbolic and post-linguistic, and can be applied to image-schematic representation: the cognitive grounding system for abstract concepts. Iconic gestures on the other hand are proto-mimetic and non-linguistic, and can be applied to mimetic-schematic representation: cognitive maps of meaningful gestures that occur before the emergence of language, and are therefore concrete representations. Co-speech gestures add a layer of communicative intent and convey information which supplement the speech acts they accompany. Indeed, it has been repeatedly shown that the brain processes and integrates speech and co-speech gesture in semantic memory, contributing to the overall mutual understanding between communicative participants (Gentilucci, et. al., 2013; Holler, et. al., 2014, 2015; Goldin-Meadow, 2016; Chui, 2014). Further evidence for co-speech gesture in semantic memory can be seen in its facilitation of both vocabulary learning and the fact that gesture aids in the retrieval of linguistic concepts.

According to Pulvermuller (2013), a word or gesture does not become meaningful unless semantic memory can make the necessary connections between the word or gesture and an existing schema. For this to happen, the word or gesture must be seen as intentionally communicative. The communicative intentions of one individual
and the specific meaning that is gained from these intentions depends on how the acts are perceived by the observer (Andrén, 2013; Holler, et. al., 2015). A study done by Osaka, et. al. (2012) attempted to view separately the contributions of the visual and intentional areas of the brain during action observation. In their words, “as intentional stance increased, the portion of the social brain involving the representation of an agent's intentional actions became more activated.” (p. 6). That is, the more the observer attributes intention to an observed agent's actions, the more active became the observer's neural representation of said actions. So, when speech and gesture are not intentionally coupled, there is less integration due to the fact that communication must be intentional in order to carry meaning.

Communicative intent and goal-oriented behaviour are mediated by mutual gaze, which literally communicates intent, whether it is meant to stand for something or not. In the case of co-speech gestures, the eye-gaze of the communicator affects the way a communicative recipient integrates the gestures with speech (Zlatev, 2013; Holler, et. al., 2014, 2015). In a study carried out by Holler, et. al. (2014), one participant was to tell a story to two others, either with or without co-speech gestures. The goal was to provide evidence for the claim that gestures aid in the comprehension of speech, while at the same time factoring in the effect of eye-gaze. In the no-gesture condition, participants who were not being looked at during communication processed speech slower than those that were being looked at. In the condition where gesture was used to supplement speech, participants who were not being looked at processed speech at the same level as the ones who were being looked at. Furthermore, despite eye-gaze alternation, in the gesture condition speakers kept their gestures in front of their bodies, allowing global access to the communicative intention.

When a language is multimodal, it uses more than one modality to create a coherent meaning. I have mentioned the manual, the facial and the vocal modalities as being equally important in communication as is speech. When one modality is lost, the others suffer, and the same goes for the understanding of a communicative act. “Gestures influence the understanding of speech, just as speech influences the understanding of gestures.” (p. 47, Corballis, 2014). Similar to a non-human primate's 'honest signals', some researchers believe that facial and manual co-speech gestures are one's thoughts being expressed in a concrete way (Quaeghebeur, 2013; Trevarthen, 2013). Thus, 'dishonest' signals, or when the body gives away one's spoken lies, often emerge when speech and gesture are incongruent. For example, those who suffer from
Bell's palsy, which causes paralysis of the face, have difficulty formulating coherent thoughts. A neurological integration of multimodal communicative acts must therefore take place in order to gain global coherence.

6. Conclusion
In this thesis, I have outlined the building blocks of the development and acquisition of language within the individual and the species. Structured language is nothing but mechanical without bodily gesture and intonation, and that these bodily intonations and gestures allow structured language to flourish. The body has an incredibly important role in communicative ability, and thus one can say that human language is multimodal. When one modality suffers, it brings the other modalities down with it, including global understanding on both the part of the speaker and the listener. This can be quite clearly seen in casual texts and e-mails, in which emoticons that convey facial expressions or sentiments are almost essential in order to convey a holistic meaning. The lack of such 'facial expressions' can cause many misunderstandings in communicative intent.

At all levels of cognition, there is a movement from more concrete modes of processing to more abstract systems of organisation. This thesis goes over the cognitive modes of communicative processing and organisation and the crucial movement from mimetic-based representations that are respective of our evolutionary ancestors, to image-schematic based representations, that are particular to today's modern-day human. Indeed, there has been much debate as to whether human language developed from gestural origins, or whether, as Chomsky debated, language evolved due to a mutation that rendered human language and non-human primate communication systems disparate. As the theories presented in this paper indicate, neither is truly a correct representation of our linguistic origins. Gesture is multifaceted and is represented and processed in both concrete and abstract ways during communication. During the earliest months of a human's linguistic development, concrete, iconic gestures are supplemented by equally concrete vocalisations. As the individual matures, concrete vocalisations show a movement into more systematic and structured speech-acts. These speech-acts become arbitrary, normative, and abstract, the gestures themselves moving from being solely iconic and concrete to becoming symbolic and abstract, supplementing speech rather than the opposite. Indeed, gesture and language are different from each other, but they are also intricately entwined.
Imitation, or mimesis, is the underlying cognitive function that facilitates social interactions, most essentially culture. It was once believed that the structure of human thought could reveal the underlying structure of culture and society, but with this thesis I have shown the opposite to be true. Our closest relatives, the non-human primates, have culture. They use tools and recognise kinship hierarchies. The structure of human culture is indicative of human thought. The skills needed to develop tools and other cultural artefacts are the skills that are crucial to the development of grammar, syntax, and goal-oriented, intentional communication. The question of language has long been about the concept of the sign and its representational qualities. Meaning is, essentially, achieved through an understanding of the sign's existence and its role in the intention to communicate something. Meaning is what allows the sign to flourish, and not the other way around.
Reference List


Handbook of Cognitive Psychology.


