STUDIA SEMIOTYCZNE—ENGLISH SUPPLEMENT, vol. XXVIII (2016), 112–149
ISSN 2451-2745
DOI: 10.26333/stsen.xxviii.05

Sylwester Orzechowski[1], **Sławomir Wacewicz**[2], **Przemysław Żywiczyński**[3]

The Problem of "Modality Transition" in Gestural Primacy Hypothesis in Language Evolution: Towards Multimodal Hypotheses

Abstract In our paper we review the gestural primacy hypotheses in language evolution, starting with the discussion of the historical advocates of this approach and concluding with the contemporary arguments, derived from empirical research in various fields of study. Assessing the strengths and weaknesses of the gestural scenarios we point to their main problem, namely their inability to account for the transition from a mainly visual to a mainly vocal modality (the so called "modality transition problem"). Subsequently, we discuss several potential solutions to this problem, and arrive at a conclusion that the most satisfying option is the multimodal perspective, which posits that language evolved as a bimodal system, with the vocal and visual modalities very closely integrated from the very early stages.

Keywords language evolution, gestural protolanguage, modality transition problem, multimodal hypotheses

[1] Institute of Psychology, Maria Curie-Sklodowska University.
E-mail: sylwesto@gmail.com
[2] Center for Language Evolution Studies, Nicolaus Copernicus University.
E-mail: wacewicz@umk.pl
[3] Center for Language Evolution Studies, Nicolaus Copernicus University.
E-mail: przemek@umk.pl

1. Introduction

1.1. Research background – the evolution of language

Despite the belief that language origins cannot be pursued scientifically (Campbell 1998; Fisiak 1985) still being present in the 20th century, the *evolution of language*, research on the phylogenetic source and development of language ability, has become a well-established, recognisable and dynamically developing field of study. According to the data from *ISI Web of Science*, the number of publications on language evolution increased tenfold from 1990 to 1999, and threefold from 2000 to 2009 in proportion to the previous decade. There are textbooks (e.g. Johansson 2004; Hurford 2007) and encyclopaedic compendia (Tallerman, Gibson, 2011) on this domain. Furthermore, the debates on evolution have become central to linguistic inquiry (Hauser, Chomsky, Fitch 2002; Pinker, Jackendoff 2005). This change has mainly been possible through significant advances made in empirical studies in recent years – particularly, in the research on communication and cognitive abilities of non-human animals, neuroscience, genetics, and computer modelling. A more thorough description of the subject comes from Fitch (2010), Tallerman and Gibson (2011); an overview in Polish comes from our previous writings (e.g. Wacewicz, 2008, 2013; Żywiczyński and Wacewicz, 2015).

1.2. Gestural primacy hypotheses in language evolution

According to gestural primacy hypotheses, sometimes referred to as gestural[4], language phylogenetically stems from a gestural form of communication conveyed in the visual channel (see 1.3.). In an obvious way, these explanations of language origin compete with vocal-auditory hypotheses that trace back the origin of language to non-linguistic vocalizations (e.g. Burling 2005; Dunbar 1998; Mithen 2005; MacNeilage 2008). Vocal hypotheses are usually based on the intuitive assumption that the development of human communication from the original to current language form was shaped entirely by the vocal modality. Although this assumption is intuitive, it should not be taken as implicit. The vocal-auditory modality is so firmly established that other proposals on language origin are hardly ever noticed (e.g. Kenneally 2007). Lack of awareness on the status of sign language – language in the full sense of the word – is a related problem (see 3.2.); Charles Hockett's

[4] Terms such as "*gestural primacy hypothesis*" (GPH); *gestural hypotheses*; *gesture-first hypotheses*; *from hand to mouth*; *language from gesture* are present in the literature; *gesture together with speech* falls under this category as well.

designed features of language (e.g. 1960) constitute a familiar and telling example, which in its initial form relates to speech and the exclusion of gestural modes of linguistic expression.

Gestural primacy scenarios have been considered by many authors, representing both speculative philosophising and contemporary science. There are a few types of gestural primacy hypotheses, differing in descriptions of gestural and vocal communication, and organisation of the postulated gestural system. These types include:

– gestural language before speech hypotheses,

– gestural protolanguage hypotheses.

Gestural hypotheses also peripherally embrace:

– *gesture together with speech hypotheses.*

Gestural language prior to spoken language hypotheses (e.g. Corballis 2002; Stokoe 2001) assume the stage of developed gestural language before speech emerged. Although these authors do not rule out the role of vocalisation in language development, they highlight that its function was limited to transmission of nonverbal information such as emotions.

Gestural protolanguage hypotheses in turn assume that protolinguistic communication mainly relied on gestures functioning as simple, syntaxless signs. Gestural primacy hypotheses come in two varieties:

– *synthetic gestural protolanguage* – gestures stand for words and represent referents (objects or actions) that can be combined into short strings with compositional content (the meaning of a whole stems from the meaning of components), but lack the syntactic or morphological structure (e.g. Hewes 1973);

– *holistic gestural protolanguage* – individual gestures are equivalent to whole utterances, representing complex thoughts or situations, e.g. "I am hungry" (Arbib, 2005).

Gesture together with speech hypotheses, based on the close link between speech and gesticulation during linguistic expression, assume that language development always comprised both modalities: vocal and visual (e.g. Goldin-Meadow 2011; McNeill 2012; Kendon 1991). We discuss this problem in detail in section 5.

In spite of the differences between these approaches, they are viewed jointly. The gestural component unites all the aforementioned views in the sense that they oppose the intuitive assumption about the dominance of the vocal-auditory channel in the evolutionary history of language. Our paper does not discuss the arguments for gestural hypotheses extensively (for more see Corballis 2002; Armstrong, Wilcox 2007; Fitch 2010) but reviews them briefly in sections 2.2. and 2.3. Our goal is to make an attempt to solve the core problem of most gestural hypotheses – the problem of modality transition described in section 3, which could be described as follows: *if language originated as a system of gestural expression, how can we account for its transition to the current, mainly vocal form*?

1.3. Gestures – definition

A straightforward and theoretically neutral definition of gesture is problematic due to the multiple meanings of the term – both colloquial and technical. Broadly speaking, gesture comprises every expressive movement – expressing an emotion or thought, performed by means of any part of the body (*Oxford English Dictionary*, after: Kendon 2004), including the face and eyes. Kendon (2004) proposes that only those actions that can be interpreted as volitional and intentionally expressive rather than serving any other purpose (e.g. pragmatic) should be termed as gesture. In accordance with this strict approach, the term gesture is reserved mainly for idiosyncratic and spontaneous hand and arm movements synchronised with speech (McNeill 1992). Some authors ascribe this term to instrumental actions (oriented towards physical objects) such as hand grasping (see Fogassi, Ferrari 2004). Interestingly, the articulatory movements are sometimes classified as gestures (described in more detail in section 4.5.).

Gestures do not solely belong to the domain of human communication. Importantly for language evolution studies, nonhuman great apes use this form of communication too (DeWaal, Pollick 2011; Pika, Liebal, Call, Tomasello 2005; Pollick, DeWaal 2007; Tomasello 2008), as well as more distantly related to us Old World Monkeys (Maestripieri 2007; Meguerditchian, Cochet, Vauclair 2011). Below, we compare two different perspectives on defining gesture: the interpersonal communication perspective and the primatological perspective.

1.3.1. Gestures in interpersonal communication

Gestures are an integral part of the human communication system. Due to a number of forms they assume and functions they perform in message transfer (Goldin-Meadow 2003), it is not easy to classify them into distinct categories. One of the most influential descriptions of gestural behaviours – *gesture continuum* – comes from McNeill (1992, 2005, 2012). Within this idea, gestures are placed on the continuum of behaviours as follows:

gesticulation–language-slotted gestures–pantomime–emblems/deictics–sign languages

The continuum is arranged with respect to three criteria (from the left to right side of the continuum): 1) the role of speech decreases; 2) the presence of linguistic features increases; 3) the level of conventionalization increases as well.

The term *gesticulation* refers to the hand and arm movements accompanying speech (Kendon 2004). However, although such movements are closely connected with narration, they do not show any linguistic systematicity. Their form stems from an ongoing coupling with meanings expressed by words and is of a spontaneous character. Gesticulation is not a simple category with various authors differently classifying gesticulations types (Ekman, Friesen 1969; Krauss, Chen, Gottesman 2000); again, McNeill's is the most influential one (McNeill 1992) and distinguishes the following categories:

- *iconics* – resemble the semantic content of an utterance by representing concrete objects or actions;
- *metaphorics* – similar to iconic gestures; however, they represent more abstract concepts or ideas;
- *beats* – the form of these gestures is the same (usually rhythmical movements of the arm or hand up and down, front and back, or left to right); they lack semantic reference to the content of speech but are synchronised with its rhythm;
- *deictics* – their main function is pointing to objects or activities, not present in any physical way, in the context of an ongoing communicative situation. Due to their structure and function, they may constitute a separate category, but along with emblems, they may also be subsumed under gesticulation.

Language-slotted gestures bear similarity to gesticulation but differ in the syntagmatic relationship they have with words. Let's look at the following example: "the weather was good, but the food [hand wave]." The gesture completes the sentence and fills the syntactic blank slot that arises as a result of the omission of the predicate. *Pantomime*, on the other hand, is of a completely different character – it cannot be accompanied by speech. Here, objects and actions are represented by means of gesture sequences. A string of gestures is an example of pantomime: calling someone by a hand movement, holding a finger against the lips to mean "keep silent" and pointing towards a place where joint attention is directed. Although pantomime is characterised by a sequential organisation of signs (in the manner indicated above), it does not show the properties of syntax. Next, the main function of *emblems* is to replace individual words. In their gestural repertoire, each culture or society has a set of emblems that are characterized by arbitrariness and conventionality (special rules of conduct require that emblems be produced in a particular way – in analogy to the phonological rules in language), intentionality[5], and cultural transmission. An example of the Polish emblem is the "OK" gesture, performed by raising the arm and putting the tip of the thumb and of the index finger against each other, with the rest of the fingers directed upwards. Emblems are reminiscent of linguistic signs, but they appear as single signals transferring short messages. *Signs of sign languages* occupy the right extreme of McNeill's gesture continuum. Sign languages have all the features of language systems, and the crucial difference between them and spoken languages is the modality – vocal-auditory for speech and motor-visual for gesture (see section 3.2.) It should be further stressed that despite common intuition, signs of sign languages constitute a system that is completely different from gesticulation not only in terms of its formal and communicative manifestation but also cortical localisation.

The above account situates gesture in the context of interpersonal communication, related to specific cognitive abilities, social structure and the presence of language. Defining gestures in different contexts, e.g. in primate communication, requires consideration of different conditions. Below, we present selected definitions and typologies of gesture derived from primatological research.

[5] Here and further in the text we use the term intentional in the psychological sense: "deliberate", "resulting from the intention of the subject" – not in the philosophical sense: "having intentional content", "on something".

1.3.2. Gesture in the communication of nonhuman primates

Although primates use various means of communication, their gesture usage has drawn particular interest of researchers. Gestures differ from other communicative behaviours (de Waal, Pollick 2011; Pollick, de Waal 2007; Tomasello 2008) in that they involve:

– acquisition through individual learning,

– intentional use and flexibility,

– relative independence of emotional processes,

– targeting specific recipients.

Gestural communication based on ritualised behaviours is present mainly in nonhuman apes, which suggests the late phylogenetic emergence of this form of communication (de Waal, Pollick 2011).

Comparative studies on gestural communication of nonhuman apes (bonobos, *Pan paniscus*; chimpanzees, *Pan troglodytes*; gorillas, *Gorilla gorilla*; orangutans, *Pongo pygmaeus*) showed that their gestures resemble (to an extent) gestures of infants and toddlers who are right at the beginning of language acquisition (Pika, Liebal 2006). Similarities concern the diversification of gestural repertoire and its intentional use in both groups. Differences are connected with the nature of gestures and the way they can be used: the majority of gestures in nonhuman apes are of dyadic character – the sender drawing the receiver's attention expresses a wish (the imperative gesture use); children, on the other hand, can use gestures triadically – they direct the receiver's attention to an external event or object in order to share attention or comment upon the shared object of attention (the declarative gesture use). These differences stem from a different cognitive bias (understood as "social-cognitive abilities"; see Tomasello 2008), in human and nonhuman apes. Furthermore, they shed light on language origin problems and the nature of interpersonal communication.

Primatologists studying nonhuman apes' communicative behaviour developed a few gesture classifications. DeWaal and Pollick (2011 as well as Pollick, deWaal 2007) propose to limit the term gesture to manual movements, which exhibit features different from other communicative movements (see gesture characteristics, above). Tomasello (2008) characterises gesture broadly as intentional communicative behaviour executed in the visual modality, mainly body posture, facial expression and manual gestures, characterized

by flexibility, and acquired by means of ontogenetic ritualisation. Pika (2008) proposes a slightly different definition. She explains gestures as expressive movements of the limbs, head and the whole body, which:

- target specific recipients,
- are mechanically ineffective – such as e.g. pushing someone – which differentiates gesture from instrumental movements,[6]
- provoke specific reactions,
- are performed intentionally (deliberately).

Intentionality is a crucial criterion in the aforementioned definitions. It enables the differentiation of gestural behaviours from other communicative behaviours of intrinsic and hence non-intentional character, which are found in animal communication (Tinbergen 1951). The intentionality of gestures is established by the following criteria:

- *relative context-independence* – the same gesture is used to attain different communicative goals, and different communicative goals are attained by the same gesture (de Waal, Pollick 2011; Tomasello 2008),
- *audience-checking*,
- *response-waiting*, and
- *persistence*.

As the problem of sensory modality is key to our argument, we will adopt Simone Pika's (2008) gesture typology, which is largely modality-based; she distinguishes the following gesture types:

- *auditory gestures* – accompanied by sound production, e.g. hand clapping or chest-beating in gorillas;
- *tactile gestures* – involve physical contact between the producer and receiver, e.g. a directed scratch in chimpanzees; the first referential gesture discovered in wild nonhuman apes (Pika, Mitani 2009);

[6] The difficulty arises as the criterion complicates the recognition of an important gesture class – touch gestures – which require some mechanical influence on the receiver.

– *visual gestures* – engage only the visual modality.

In our paper, we adopt a broad definition of gestures. We start with an intuitive understanding of gestures as intentional, communicative and hence non-instrumental hand and arm movements, but we extend the definition to the majority of bodily signals operating in the visual modality.

The prototypical examples of gestures are arm, hand and finger movements performed in order to transfer information. More peripheral examples embrace a variety of visual signals, such as proxemic behaviours, body postures and positions, consciously produced facial expressions or even gaze patterns. Speech-related articulatory movements, although they may constitute a continuum with the aforementioned movements, especially facial expressions, are included in the gesture repertoire only if they can be perceived visually.

2. Arguments in favour of gesture primacy hypotheses

The first speculations on the role of gestures in language evolution had appeared long before the idea became an object of science. It was not until the 1970s that the gesture primacy hypothesis was based on more extensive empirical foundations. Currently, its various versions jointly constitute the most influential position in the discussion on the phylogeny of language. In the present section we present an outline of the history of gestural hypotheses and sum up the most important evidence supporting these positions.

2.1. Gesture and language origin – a brief historical background

For many centuries – since antiquity until the 18th century – it was commonly held that gestures constitute a natural form of human communication – an autonomous and universal language. The supporters of this idea included: Roman rhetorician Quitilianus, and later, Giovanni Bonifacio (1547–1645) and John Bulwer (1606–1656). Similarly, the inventors of the first sign language systems, Charles-Michel de l'Épée (1712–1789) and Roch-Ambroise Cucurron Sicard (1742–1822), believed that gestural communication is a perfect basis for the creation of a universal language independent of speech. Diderot (1713–1784) argued that gestures used by the deaf provide a direct insight into cognitive processes as they are free from the distorting influence of convention and tradition.

The problem of gestures was also present in speculations on language origin. Condillac (1715–1780) assumed that in the time before people learned to control speech, they had communicated by means of gestures and body

movements. Gestural primacy was also assumed by Giambattista Vico (1688–1744), for whom gestures were a perfect way of representing visual experience. Edward Tylor (1832–1917) concentrated on how thought can be expressed, including through the means of gestural languages, pictograms and writing systems, which led him to the conclusion that studies on gesture and pictograms may be helpful in understanding the origin of language. Wilhelm Wundt (1832–1920) also sought to trace back the source of language to expressive movements, the patterns of which are dependent on emotional experience. Interest in gestural scenarios abated in the first half of the 20th century and arose anew in the beginning of the 1970s.

2.2. Hewes's position and the revival of interest in gesture in language evolution

The modern hypothesis of gestural primacy was formulated by American anthropologist Gordon W. Hewes in the article *Primate Communication and the Gestural Origin of Language* (1973). The article is a synthesis of data coming from various fields of research and lays out a range of arguments supporting the gestural scenario of language origin. Some of the arguments postulated by Hewes are still used in the ongoing debate on language evolution (see Corballis 2002; Tomasello 2008), others have either been updated (in line with the newest research) or rejected. In the following section, we look at the most important arguments from Hewes's original presentation.

Hewes's main argument pertained to the *relative success in teaching nonhuman apes a version of sign language* (Gardner and Gardner, 1969), which contrasted with the complete failure of many attempts to teach them spoken language. It can therefore be assumed that early hominins whose cognitive abilities must have been much like those of modern nonhuman apes were capable of creating a gestural protolanguage.[7] Although early hominins, just as the other primate taxa, used vocal communication, Hewes suggests that it could not have been the starting point for the development of linguistic communication, which is compositional, propositional, and relies on conventional-arbitrary signs, etc. The main obstacle was the lack of volitional control over their voice. Based on the research of the time, Hewes

[7] Hewes was the first to use the term *protolanguage* referring to the transition period between nonlinguistic communication of apes and fully linguistic communication. The current understanding of the term which in its simplified form means language deprived of grammar – in the sense of morphological rules or syntax – was promoted by Derek Bickerton (1990).

noticed that vocal reactions of nonhuman apes are triggered by emotional stimuli to which they relate; moreover, such vocalisations do not have an addressee – they are nonselective – and can be elicited in the absence of any other animal. These facts were viewed against the gestural communication of nonhuman apes, which is characterised by volition, flexibility and is based on higher cognitive processes (importantly, the available results of the current research – although they do not question the aforementioned qualitative difference – point to a much more complex nature of vocal communication in nonhuman apes than was previously thought[8]).

Hewes relied on the research indicating that a *human vocal tract is a relatively late adaptation*, found in *Homo sapiens* only (c.f. section 3.1.). At the same time, he believed that the species preceding *Homo sapiens* had the abilities that required at least an elementary form of language: usage of fire, tool manufacture and big-game group hunting. Hewes assumed that gestural protolanguage was enough to sustain a culture organized around such activities.

Hewes (1973, 1981, 1996) addressed a few crucial problems that continue to be discussed in the ongoing debate on the roles of gesture in language evolution. He underlined a possible role of the deictic gesture in the early stages of language development, he noticed the phenomenon of gestural imitation as a potential way of establishing a linguistic sign, he also pointed out the significance of lateralisation and movement control in shaping gestural language. Another interesting argument articulated by him focused on the depigmentation of the inner part of the palm in non-white individuals – Hewes (1996) suggested that this property may serve to increase hand-visibility and hence, it might have emerged as an adaptation for gestural communication.

Lastly, Hewes articulated the problem of "modality change", and proposed his own solutions based on the idea of mouth gestures and sound symbolism as evolutionary mechanisms of transition from gesture to speech[9] (see also section 4.5.1.).

[8] It is characterised by *functional reference*, *audience effect*, *productivity* (without compositionality) and *tactical deception* – overview of the new data e.g. in Slocombe (2011).

[9] "A plausible theory of the primacy of gestural language over speech must, as has already been noted, account for its general replacement by spoken language" (Hewes 1996: 587).

2.3. Contemporary gestural hypotheses

Currently, gestural hypotheses are still being developed and hold an important position in the debate on the phylogeny of language. In the following section we present a few lines of evidence supporting the gestural hypotheses that appeared after Hewes. They point to the iconic potential of gestures as the most intuitive form of linguistic expression, new findings on brain functions, and *mimesis* – the uniquely human form of imitation.

2.3.1. Iconicity of gestures

William Stokoe, one of the pioneers of modern research on sign language, demonstrated that the expressive potential of sign languages is equal to that of spoken languages and proposed solutions to key problems of language evolution (Stokoe 1960). The first of these relates to the emergence of arbitrary speech sound and meanings, which can be addressed with reference to a gestural stage in language evolution, whereby the iconicity of gesture (the resemblance of the visual form to content), creates a bridge between sound and its referent. The other problem is connected with the origin of grammar. According to Stokoe (1991; later also Armstrong, Stokoe, Wilcox 1995), the iconicity of a gesture allows us to represent simultaneously the activity, and the agent that performs the action as well as the patient that is affected by the action. In this respect a single gesture can represent an action as a complex whole: the hands and arms function as a prototypical noun, their action is a prototypical verb – collectively, they create a prototypical sentence (Armstrong, Wilcox 2007). The spatial nature of gesture further facilitates an intuitively clear visualisation of semantic roles (the hand hits the hand or draws a path) and spatiotemporal relations. It is the reference to the gesture stage in the evolution of language that facilitates the explanation of the origin of the arbitrariness of speech and the origin of grammar.

On a par with theoretical arguments supporting this line of reasoning there are also interesting empirical data, especially from the experiments conducted by Susan Goldin-Meadow's research team. In one of the experiments, (Goldin-Meadow et al. 2008) the respondents were describing simple activities by means of speech, gesture and pictures. Verbal descriptions were compatible with the structure of the sentence of the respondents' native language. However, the gestural or pictorial descriptions, independent of the native language of the respondents, were characterised by a stable order: "actor – patient – act", congruent with the SOV word order. Based on that it can be purported that gestural communication presents a somewhat "nat-

ural" or "original" form of the mental representation of event structure. In another experiment (Fay, Arbib, Garrod 2013), the respondents, by means of gestures or nonverbal vocalisations, communicated "meanings" from the finite repertoire of meanings – emotions, objects, or actions. Gestural communication proved more effective than the vocal one. The authors of the research considered it a piece of evidence in favour of the visual modality for motivated expressions, and stated that it might have been useful in the initial stages of language development.

It should be noted, however, that the status of iconicity as a cognitive facilitator inhibiting the transition to symbolic communication is far-fetched. For instance, in ontogenesis, the iconic representations do not seem privileged over arbitrary ones, e.g. children do not acquire them easier or faster (Tomasello 2008: 147). Also with respect to the cerebral realisation, the processing of iconic gestures and symbols are dissociated (Niederhut 2012).

2.3.2. Handedness and lateralization

Handedness and lateralisation were for a certain period taken as supporting the gestural scenario of language origin. In the majority of people, the left hemisphere of the brain is responsible for both language processing tasks (it is here where "language areas" are located) and motor control of the dominant hand (90% of all people are right-handed). The sources of this correlation are not entirely clear, but the available data show explicitly the systematic character of the correlation, e.g. the degree of language processing in the left hemisphere is directly proportional to the level of preference towards right-handedness (Knecht et al. 2000). There are also hypotheses on language and lateralisation. Hewes (1973) assumed that lateralisation for precise movement control, and hence gesticulation, emerged before speech. According to Corballis (2003), representations of communicative movements of hands and arms were gradually absorbed by vocalization governed by the left hemisphere of the brain. Furthermore, according to the throwing hypothesis proposed by Calvin (1982, 1983; see also Calvin, Bickerton 2000), one of the first expressions of the lateralisation of brain functions and handedness was the action of throwing stones to hunt small game. Well-aimed throws require the calculation of a projectile's trajectory and creation of an appropriate motor plan that will coordinate movements of the body parts (including fingers, wrist, arm and torso). Calvin stated that one-hand throwing led to the selection of a dedicated neural circuit in the brain that facilitated such computations. The circuit was then exapted for tool

production, gesticulation and eventually for articulated speech and syntax.

Currently, the relevance of handedness and lateralisation as evidence supporting gestural hypotheses is decreasing. First, the newest research challenges the position that handedness at the population level is a solely human feature. Although the data are ambiguous, it seems that our species differs from others only in the scale and systematicity of the described phenomena (e.g. Cashmore, Uomini, Chapelain 2008). If we look at nonhuman great apes, some studies (e.g. Harrison 2008) do not confirm any preference at the level of population, while others (e.g. Hopkins 2006) note its presence, most often right-handedness, at least in the case of some species and some activities. Similar data exist for nonhuman apes – for example baboons show the preference to perform communicative gestures, but not instrumental ones, with their right hand (Meguerditchian Cochet, Vauclair 2011). Furthermore, the left hemisphere controls vocal communication in many species including those distantly related to humans, e.g. frogs and birds (Corballis 2003). These observations undermine the assumption that handedness is uniquely human related selective pressures from the hominin history, such as the need for precise motor control necessary to manufacture tools. Brain neuroimaging proved that the term "left lateralization for language" is far-fetched, and the functions responsible for language processing engage various areas all over the cortex (e.g. Deacon 1997; Lieberman 2003).

2.3.3. Broca's area and mirror neurons

Neurocognitive research provided a few interesting clues leading to gestural hypotheses: they concern the revision of understanding the function of Broca's area and the discovery of mirror neurons. The role of Broca's area, believed to be primarily connected with speech production, change in the light of the newest research (Fadiga, Craighero, D'Ausilio 2009). It was proved that this area is engaged in comprehending language, performing and observing manual activities, performing and listening to music and representing abstract hierarchical structures. On this basis, some authors (Fadiga, Craighero, D'Ausilio 2009) conclude that Broca's area is responsible for detection and representation of complex hierarchical/syntactic relationships, irrespective of modality and use (i.e. whether it is used for production or reception). It is hypothesised that the evolutionary foundation of these abilities stems from the motor function connected with performing activities and – thanks to mirror neurons – their comprehension. This data suggest the equation of the visual and the vocal modality as potential ways of the transmission

linguistic information, concurrently granting phylogenetic primacy to the motor system.

An interesting perspective on language evolution was proposed by Rizzolatti and Arbib (1998): mirror neurons – a set of nerve cells that fire during both performing an action and observing how another is performing the same action. Mirror neurons were initially discovered in the rostral part of the ventral premotor cortex (area F5) of macaques, which is situated roughly around Broca's area in the human brain (Rizzolatti et al. 1996). Later studies confirmed the existence of mirror neurons in humans (Iacoboni et al., 1999). The hypothesis of mirror neurons identifies the neural mechanism responsible for equivalence and reversibility of the sender and receiver roles in a communicative act (parity requirement) as a preliminary and essential condition of information transfer. A nonhuman primate seeing its kin (or human) reaching for a peanut, understands the meaning of this action due to a copy of motor representation of the event in their own brain. Thereby, a mental link is created between the "sender" (actor) and the "receiver" (observer), which represents the same information simultaneously.

Arbib (2002, 2005, 2012) modified the initial idea, emphasising, along with the function of mirror neurons, the role of imitation and volitional control over communicative movements. The gradual model of language phylogeny was established: the first three stages involve all primates until the emergence of the common ancestor of the chimpanzee and Homo sapiens, the consecutive stages relate to changes after the split of the two taxa. The model is presented by Arbib in the following way (2005):

S1: grasping,

S2: mirror system for grasping,

S3: simple imitation (chimpanzees only),

S4: complex imitation (after the homo-chimpanzee line split),

S5: proto-sign (key change leading to the emergence of an open repertoire of signs),

S6: proto-speech (key change leading to the motor control over voice),

S7: fully fledged language.

Although manual activities and the mirror system still constitute the basis of the model, the key feature granted to primates is the use of pantomimic

imitation comprised of interim proto-signs representing whole situations and activities (holistic protolanguage). Furthermore, due to the extended motor control of the tongue and larynx, the vocal modality and arbitrary symbol-based communication might be gradually incorporated.

2.3.4. Mimesis and pantomime

Another influential line that draws on imitation stems from the mimetic concept of language evolution, initially developed by Merlin Donald (1991, 2001). According to Donald, the fundamental difference between nonhuman primates' and humans' cognitive processes is related to mimesis – the ability for conscious, self-initiated, intentional representations lacking linguistic character. The mimetic ability enables us to remember motor schemas for such activities as jumping, throwing or dancing, as *activities that are detached from their subject*. Such representations do not have to be activated as a direct stimulus reaction – the object may consciously recall them from memory at any time (*autocueing*). Therefore, one may perform an activity at any time – be it an actual dance movement or just a dance-based exercise. Such a sequence of movements may also be recalled only from the working memory – where one imagines it "off-line" – as an action plan. The subject capable of mimetic representation can separate movement sequences (e.g. throwing a stone) observed in others, remember them, imagine them, and re-enact them on one's own. In this way, mimesis underpins imitation. An activity represented mimetically – be it the imitation of a stone throw – is identified as the same by both the actor and the observer; hence, it can be used for communication. It should be noted that such mimetic signs are neither arbitrary, nor conventional, nor compositional, and their nature is entirely corporeal. Furthermore, although mimesis is a multimodal ability, the visual domain is crucial in this case.

Jordan Zlatev (e.g. 2008)[10] offers the most detailed explanation of Donald's concept. He puts forward a hierarchy of mimetic abilities arranged into the stages of their phylogenetic development:

– *proto-mimesis* – based on imitation taken from external observations (exteroception, e.g. vision) of one's own body movements (proprioception, e.g. kinaesthesia), present in nonhuman apes and ontogenetically in newborn babies; it manifests itself through such activities as eye contact or simple coordination of behaviour;

[10] The mimetic standpoint of Zlatev's (e.g. 2008, 2013) is nonetheless closer to multimodal theories (see section 5), than *stricte* gestural ones.

– *dyadic mimesis* – based on volitional representation, present in its basic form in nonhuman apes; it underpins imitation or imagination, including representation of the future;

– *triadic mimesis* – based on communicative intention, present in humans and some enculturated nonhuman apes; it manifests itself e.g. through purposeful iconic gestures or declarative pointing;

– *post-mimesis 1* – based on normativeness and convention; present in humans (and in nonhuman apes taught symbolic communication); it manifests itself through symbolic communication;

– *post-mimesis 2* – based on systematic and compositional usage of symbols in both communication and internal thought processes; only present in humans; it underpins the language ability and the ability to understand false premises.

Daniel Hutto (2008) espouses a similar position – the initial form of communication was, according to him, a mimetic re-enactment of events not segmented into any meaningful units, but presenting a holistic, bodily representation. Hutto assumes that regular re-enactments of events might have had an important social function, establishing a basis for customs, strengthening bonds and gradually replacing grooming. Michael Tomasello (2008) acknowledges pantomime and pointing as the natural and first specifically human forms of communication, representing a transition phase from nonhuman apes' communication to conventionalised language. Their basis (just as further-developed conventional languages) are founded mainly on the exclusively human form of sociality and motivation: shared intentionality based on the recursive mindreading and cooperative communication.

2.3.5. Further arguments

Wacewicz and Żywiczyński (2008) have underscored greater secrecy of communication in the visual channel, which is suited to the transfer of a message to a selected addressee rather than all individuals close-by. Evolutionary logic suggests that such selectivity might have been used at the early stages of the development of communication, when it was an evolutionarily unstable system – it lacked propositional content characteristic of language but was highly manipulative – as is typical of nonhuman animals. Some recent field experiments on chimpanzees resorting to the use of gesture in conditions requiring secrecy (Hobaiter, Byrne 2012) confirm this idea.

An alternative approach to the role of gestures in the evolution of language comes from David McNeill (2012). In his view, theories assuming gestural primacy are not capable of explaining satisfactorily the deep and multilevel integration between gesture and language. These ideas diminish the role of gestures in embodying language in favour of speech. Assuming that gestures are an integral part of speech, not only its complement, McNeill proposes that both types of semiosis (global and discrete) stem from the *dynamic units of online verbal thinking* which he terms *Growth Points* (GPs). They contain ideas simultaneously expressed in gesture and speech at the same time. In this way, ideas are expressed by means of imagistic and linguistic codes. According to McNeill, the acquisition of this ability is a critical moment in the cognitive development of humankind, leading to the emergence of language. The cognitive interface, which formed the link between thought, language and gesture from the very beginning, is central to the idea. We comment upon this proposal extensively in section 5.

3. The problem of transition to speech

The key problem of modality transition can be summarised in the following way: if language emerged as a largely gestural phenomenon, how can we explain its transition to the current, mainly vocal form? Although the transition from the hypothetical gestural language to the current – mostly vocal – communication system was already acknowledged by early authors as a difficulty (Hewes 1973), the most extensive criticism comes from Fitch (2010), who stresses two points: – *the scale of anatomical and neural adaptations of humans for spoken language*, – *the completeness and functionality of the currently existing sign languages.*

3.1. ***Homo sapiens's*** **adaptations to speech**

First of all, humans differ significantly from their closest relatives – nonhuman great apes – in terms of the anatomical structure and motor control of the vocal tract. The human species-specific features include:

- a descended larynx,
- a lack of air sacks,
- a better innervation of the thoracic muscles,
- advanced ability for vocal imitation.

For some of these differences alternative explanations cannot be conclusively ruled out. For example, the lowered larynx may be a side effect of the erected posture or the reconfiguration of the face, or as the result of pressures for the enlargement of body size during vocalisation (for discussion see Fitch 2010); the loss of air sacks may be the result of the proneness of this organ to infection; the better innervation of the chest may be an adaptation for breath control during physical effort; the ability for vocal imitation may be linked to musicality. Still, the most convincing interpretation is that all these changes constitute adaptations for articulated speech.

To sum up, from the perspective of researchers such as Fitch (2010), the extent of speech adaptations constitutes a crucial argument in favour of its early development, i.e. in hominins preceding *Homo sapiens*. At the same time, it becomes an argument against the gestural hypotheses only when we assume a "late" emergence of language. The facts relating to speech adaptations may be made compatible with the gestural hypotheses by assuming an "early" emergence of language based on even earlier visual protocommunication (e.g. a gestural protolanguage in *Homo erectus*), with further gradual development of articulated speech. An early language emergence is compatible with the view supported by new empirical data. For example, recent studies (Johansson 2012; Dediu, Levinson 2013) conclude that the reconstructions of the Neanderthal genome, anatomy (also of the vocal tract), and ontogeny, while not directly confirming the presence of language in this species, are consistent with such a possibility. These new findings make a recent language emergence unlikely, but are consistent with the gestural scenarios, provided that the latter assume a sufficiently large timescale.

3.2. Sign languages as fully-fledged languages

A much greater problem comes from the completeness and functionality of the existing sign languages. Although it is not intuitive, this argument presents a serious challenge to gestural primacy hypotheses. To understand why this is the case, it is crucial to take a closer look at the nature of sign languages.

Sign languages are fully-fledged languages, equal in expressiveness and sophistication to spoken languages (Stokoe 1960; Stokoe et al. 1965; Emmorey 2002). It concerns both the properties of the code, the level of brain description, ontogenetic development (acquisition by children) and historical development (emergence and development of the language system itself), and most importantly, functionality. Since their status as such is officially recog-

nised by linguists, sign languages, similarly to spoken ones, are present in the databases of world languages e.g. *Ethnologue* or *World Atlas of Linguistic Structures*.[11]

Sign languages, similarly to spoken ones, use conventional and arbitrary signs that can express any abstract or metaphorical concept – an example of which is sign language poetry (e.g. Sutton-Spence 2005). Similarly to speech, they have a combinatorial structure at the level of morphology, syntax and even "phonology"; hence, the *duality of patterning* is present as well.

The cerebral areas processing sign language are largely identical with those processing speech (Corina et al. 1992), and their specialisation seems to relate to language processing in general – independently of modality – rather than non-linguistic aspects such as motor control of articulators (Emmorey 2002). Thus, these centres are different from those responsible for spontaneous gesticulation. For this reason, damage to the language areas in deaf signers results in symptoms typical of aphasia, including its language-specific nature, while the ability for gesturing (e.g. Bellugi, Klima 2001; Hickok et al. 1996) or pantomime (Emmorey 2002) is preserved.

Acquisition of a sign language takes a similar time span to that of a spoken language, and is characterised by the same stages of development and the same phenomena such as babbling (its manual equivalent) or hyper-regularisation of grammatical forms; furthermore, children do not show any preferences towards either spoken or sign language while exposed to both of them in equal measure (Petitto, Marentette 1991).

The recent example of ISN (*Idioma de Signos Nicarag uense*), the Nicaraguan sign language, illustrates the *de novo* emergence of a complete language system – with the stages of pidginisation and creaolisation comparable to those in spoken languages (Kegl et al. 1999).

Most importantly, sign language researchers have shown that in terms of their communicative potential and efficiency, sign languages are equivalent to speech (Stokoe 1960; Emmorey 2002).

Despite the aforementioned characteristics and equality of signs and speech, human verbal communication is predominantly vocal, and sign languages generally perform only secondary or auxiliary functions.[12] Such

[11] *Ethnologue*: ⟨`http://www.ethnologue.com/`⟩, *World Atlas of Linguistic Structures*: ⟨`http://wals.info/`⟩.

[12] Usage of sign language signs or gestures as a replacement for speech is usually motivated by religious causes, as in case of some medieval monastic orders; cultural causes, as in case of the Warlpiri – an indigenous people of Australia's signs; or *practical* causes, as in the case of South African hunter-gatherers San people during hunting.

populations as in Al-Sajid (Israel), Adamorobe (Ghana), Kata Kolock (Bali) or Jukatana (Mexico) are exceptions, because vocal communication is impeded by the incidence of hearing impairments. In all the remaining societies the primary form of linguistic communication is speech. This fact implies that – if we accept the gestural hypotheses with their focus on the fact that language originated in the visual modality – we must account for a change of the dominant modality – a change difficult to explain when analysed more closely. Fitch poses a question about selective pressures and mechanisms of such a change;[13] many researchers (e.g. Burling 2005; Corballis 2003; Kendon 1991, 2008; MacNeilage 2008; Tallerman 2011) agree that it is the most difficult problem for gestural hypotheses.

4. Solutions

Solutions to the problem presented in section 3., although they may seem concurrent, are logical and independent from arguments opposing or favouring gestural hypotheses. Below, we propose answers to the questions posed in section 3 that are biologically plausible: we try to establish a link between the modality transition and knowledge of anatomy, evolution and brain functions.

Two general answers are possible. Firstly, we may point to potential selection pressures facilitating the development of vocal communication despite the original gestural basis. The other possibility, more interesting and discussed in section 5., questions the very problem of "modality transition". According to this proposal, the separability of visual and vocal communication is only superficial, and the evolutionary emergence of language could have been happening in both these modalities simultaneously. We would have to account not as much for a sharp transition, but a change of emphases in the extent and kind of transferred information (nonverbal, analogue, holistic vs. verbal, symbolic, combinatorial).

[13] Fitch (2010: 434): "[A] significant disadvantage of gestural models is their difficulty in explaining the virtually complete transition to vocal, spoken language in modern Homo sapiens... Whatever their virtues, models of gestural protolanguage are incomplete without a detailed and compelling model of the transition to spoken language, as most gestural proponents have recognized" (Hewes 1973; Corballis 2002; Arbib 2005); (2010: 442): "but the lack of a plausible selective force to drive signed language into vocal language remains a compelling argument against a fully gestural, and fully linguistic, protolanguage."

4.1. Traditional arguments

The overview of solutions starts with the already-existing ideas on gestural hypotheses. We find such observations interesting but insufficient to solve the problem. Therefore, we enumerate the ideas that indicate the *shortages* of visual communication, which seemingly go against gestural hypotheses:

- speech is more economical (e.g. Knight 2000) – articulatory movements need less time and energy than hands, arms and body movements;
- speech enables communication in reduced visibility or darkness (Rousseau 1775);
- voice attracts attention more effectively (Rousseau 1775);
- speech does not engage hands, which may therefore be engaged in practical tasks (work, carrying objects) during a communicative event (e.g. Carstairs-McCarthy 1996);
- speech enables the teaching of manual activities such as toolmaking (Armstrong, Wilcox 2007);
- speech acquisition starts already in foetal life, which grants a developmental advantage to this modality (Hewes 1996);
- vocal communication facilitated the monitoring of the location of a baby, which might have been important in hominins due to their hunter-gatherer lifestyle, and with lack of constant physical contact between mother and child, as is the case in other nonhuman primates (Falk 2009);
- voice is directed to everyone and not only to a specific individual (Tomasello 2008).

Fitch (2010) criticises the majority of the above-mentioned arguments. He states that it is difficult to speak of any superiority of speech over gesture in any of these aspects. Moreover, it is easy to find a balancing measure in visual communication for all the enumerated advantages of speech. Gestures are not visible in the dark, but they are visible by the firelight, and they can be used in the tactile modality, which is used by the visually impaired signers. The visual channel gains an advantage in long-distance or noisy communication; it also successfully attracts attention in these situations.

Fitch notices that although the vocal modality frees hands and arms, the visual modality frees the mouth, which was very significant in the Palaeolithic Period – the fossil data show that hominins intensively used teeth to chew hard foods and perform various mechanical operations. Furthermore, the argument concerning the energetic effectiveness is not convincing because – as Fitch points out – speech is accompanied by spontaneous gesticulation, which eventually makes this way of communication equally costly.

The arguments not mentioned by Fitch are likewise insufficient. During teaching manual activities, verbal instructions are much less effective than a demonstration or physical guidance of the learner's hands. Hewes's argument is too weak, especially in view of the developmental data on the equal pace of spoken and sign language acquisition (see section 3.2.). Falk's remark is interesting, but it does not require articulate and propositional language but just emission of any sound. Tomasello's proposition is also compelling but easy to counterpoise. The already-mentioned advantage of gestures is the secrecy of communication allowing for a more accurate choice of addressee and limited possibility of being discovered by enemies and predators.

4.2. Information duality

Susan Goldin-Meadow (e.g. 2011) noticed that the visual modality could be used to successfully transmit both *combinatorial-segmented* and mimetic (*holistic-imagery*) information. The first possibility is realised through signs of sign language – separate units of discrete and arbitrary character, which can be combined into longer compositional structures (phrases, sentences). We deal with the other possibility in the case of gesticulation or pantomime, where information is not composed of discrete units but is of holistic and imagistic nature. Goldin-Meadow notices that the vocal modality serves the transfer of discrete, combinatorial-segmented units (phonemes or morphemes) only. Instead, the ability to transfer mimetic information in vocal communication is limited to prosodic features and onomatopoeic or sound-symbolic units; therefore, their role is secondary.

In natural circumstances, for example in a conversation between two or more people, an efficient transfer of both the aforementioned information types occurs *simultaneously*. To the first approximation, gesticulating accompanies speech almost all the time (Goldin-Meadow 2003; Kendon 2004) – it also holds for telephone conversation or speech produced by visually impaired people. The visual modality may use a segmented code, but voice cannot transfer any mimetic information. Due to this fact, effective commu-

nication in a natural conversation manifests itself in speech (which encodes information in an arbitrary and segmented way) and gesticulation, which is used to transfer holistic information. Goldin-Meadow (2008) points out that this advantage of the visual modality might have paradoxically been the reason for the transfer from the hypothetical protolanguage to speech.

Erin Brown (after Zlatev 2013) puts forward a similar proposal. According to her, the vocal modality started to express symbolic code due to the fact that sound is a naturally poor vehicle for transferring motivated meanings, which in turn facilitates ascription of nonmotivated meanings. Similarly, Kendon[14] notices that gesture is a signal indispensably having a specific spatial form and location referred to as "spatial concreteness," which is not true of speech. This feature is crucial in the transfer of displaced meanings referring to the future and past, as well as to abstract concepts and ideas. Spatial concreteness of gestures may decrease their ability to express abstract meaning. Vocal signals, on the other hand, having no spatial characteristics, are easier to comprehend as abstract. A similar intuition – on higher arbitrariness and hyper-concreteness dominating the comprehension of the abstract meaning – is presented by Hewes (1973, 1996: 587) who proposes sound symbolism as a "temporary" mechanism introducing partial justification of the sign.

4.3. Acquisition of sign and spoken languages in children

Children acquire sign and spoken languages with the same ease. As it was mentioned in section 3.2., this process is reflected in the manual counterpart of babbling. Petitto and Marentette (1991) state that the ease of sign acquisition also concerns hearing children who, having the same access to speech and sign, acquire language in sign or speech at the same pace. Furthermore, they do not show a strong preference towards one or the other type: they can master language in either modality. Ontogeny provides us with arguments for seeing the transition to the vocal modality as only a surface phenomenon: it concerns Chomskyan e-languages and not language faculty (a set of biological adaptations) *per se*. Our ability to acquire language is amodal – it does not have any preference towards the modality used. This is supported by neurophysiological evidence – as we have already stated, linguistic information, regardless of its modality, is processed in roughly the same region of the brain, which concerns both sign production and sign perception.

[14] *Protolang* 2, conference speech, 19.09.2011, Toruń.

The link between the two modalities is visible in the acquisition of spoken language and single gestures. Babbling (6–8 months), for example, is accompanied by rhythmical hand movements. Before uttering their first word, babies use deictic or even iconic gestures (around 10 months). Later, they start to combine gesture and speech of the same meaning, and further, gesture and speech expressing different meanings (saying: *give* and pointing to a fruit). Both modalities are also visible in the course of cognitive development. Golin-Meadow (2003) show that at least some newly acquired concepts (conservation task, equivalence), before having a verbal representation, can first manifest themselves in gestures.

Interestingly, many experimental studies conducted mainly by Susan Goldin-Meadow, demonstrate that early gesture use predicts a wide range of linguistic achievements in later development. For example, the way an 18-month-old child uses gestures predicts their lexical repertoire and complexity of utterances in the future (Rowe, Goldin-Meadow 2009). Another example is the aforementioned speech-gesture combination: the age in which a child can use such a combination may betray the age of when the child will start to construct two-word utterances (*give apple*).

4.4. Natural connections between the hand and the mouth

The division of communication into vocal and visual is useful, but these modalities are closely connected. The interplay is visible at both the superficial level of message transfer and the deep level of cognitive and neural implementation of the appropriate systems.

The movements of the mouth and hands seem to be largely governed by a common and phylogenetically old motor control system. It is indicated by the empirical data on the neighbouring cerebral areas controlling the movements of these organs. For example, electrical simulation of area 44 (the homologue of Broca's area in humans) evokes movements of the upper limbs and lips in rhesus monkeys (Petrides et al., after Meguerditchian, Cochet, Vauclair 2011; see also Corballis 2003). The premotor cortex of macaques (F5) includes, apart from mirror neurons, neurons that activate when a graspable object is seen: coding its size in order to prepare a plan for grasping it with the hands and mouth simultaneously (Murata et al. 1997; Rizzolatti at al. 1998). Evidence for natural motor relations between the hands and mouth in people has been collected by Gentilucci and Corballis (2006). The authors describe differences in the mouth apeture and voice range when the examined subjects are uttering a syllable depending on the

size of an object they have to catch while vocalising. A similar principle is observed when the examined subjects watch the action of grasping performed by another person, which indicates the involvement of the mirror neuron system in these processes.

It is believed that a natural hand-mouth relationship is rooted in mouth feeding behaviours, which were later exaptated for linguistic operations. This might have played a role in the transformation of gestural communication – the transition from hand gesture to mouth gesture (Gentilucci, Corballis 2006). The concept of a link between the hands and mouth is long and originates from Richard Paget's "tongue gestures" proposal.[15]

4.5. Articulatory movements as a type of gesture

Motor speech perception theory gives an interesting insight into the problem of the gesture-to-speech transition (Liberman et al. 1967; Liberman, Mattingly 1985; Liberman, Whalen 2000). Accordingly, the motor system is responsible not for only speech production but also speech comprehension. The theory holds that our mental representations of phonemes are not manifested as physical sounds, but as motor programmes, and similarly phoneme recognition is possible linked to its sound value but through the reference to movement sequence of the tongue and other articulators. This idea points to the primacy of motor aspect over an acoustic aspect allowing for the classification of *speech as gesture* as a system of orofacial gestures that also produce sounds.

Conceptualising speech as gesture entails the co-occurrence of the visual and vocal component. The most obvious manifestation of this is lip-reading, which makes it possible to reproduce part of a message without access to sound (Summerfield 1992). Another proof of a deep integration of visual and auditory information is the McGurk effect (McGurk, MacDonald 1976): placing the *ga* sound to a video where the lips pronounce the *ba* sound, makes the observer hear the *da* syllable. This data unequivocally show the multimodal character of speech perception and an important role that motor-visual components play in it.

[15] According to Paget (1930, see Kendon 2011), the tongue and other articulators movements are naturally connected with hands – they can unconsciously follow hand gesticulations and perform 'tongue gestures' which might have been the base for speech sounds. Paget's contemporary commentators were critical towards this proposal (see Kendon 2011; Fitch 2010).

4.5.1. Orofacial gestures

Orofacial gestures, according to the gestural primacy hypothesis, might have initially performed a communicative function in the visual modality, with co-sounds only later acquiring a communicative meaning. Let us note that such an idea does not identify a selective pressure but is important due to the biologically realistic mechanism. The potential of orofacial gestures did not gone unnoticed by evolutionary scientists, becaming an important element of numerous scenarios, especially those of the gestural provenance:

- Hewes (e.g. 1973) indentifies *mouth gestures* as the second, next to sound symbolism, link between gesture and speech;
- Corballis (e.g. 2003) argues that similar facial gestures might have been better identified because of the co-occurring sound;
- Studdert-Kennedy (2005) assumes that mimicry of facial expressions played a crucial role in the development of control over the articulators – recurring acts of imitation led to an increasing control of various facial regions, which expanded to vocal articulators due to the link between mimicry and vocalisation;
- MacNeilage (1998, 2008), although he is not a supporter of gestural theories, acknowledges the importance of facial gestures: based on the similarities between speech and sound-producing facial gestures of nonhuman apes (smacking), the author proposes the following scenario – opening and closing of the mouth while chewing, sucking and licking, started to acquire communicative functions and took the form of facial gestures, which further transformed into syllables and phonemes;
- Meguerditchian, Cochet and Vauclair (2011) emphasise the importance of facial gestures from the perspective of neurobiological research on nonhuman apes, which reveals a deep connections between the hand and mouth, as discussed in section 4.4.;
- Orzechowski, Wacewicz and Żywiczyński (2016; Wacewicz, Żywiczyński, Orzechowski 2016), highlight the role of *auditory feedback* – the authors extend Corballis's proposal, stating that the benefits from a sound component are not limited to the receiver, but include the sender of the message for whom it is easier to discriminate certain gestures at the level of production. This position is supported by empirical

research indicating that a delayed auditory feedback or lack of it, leads to serious problems with articulation.

5. Conclusion – towards multimodal hypotheses?

There are a few crucial problems in each of the extreme positions. Theories supporting a vocal origin of language must account for the uniqueness of linguistic adaptations: nonhuman primates' vocalisations and language are separate communication systems which have little in common apart from the use of vocal modality. On the other hand, the radical gestural perspective, assuming an original gestural language, faces the problem of modality transition: from visual to vocal. Moreover, neither of the positions is able to provide a convincing explanation for a deep integration of gesticulation and speech. A multimodal perspective facilitates the understanding of these problems: gesture and speech are various realisations of the same, common system. In consequence, we can adopt a vision of language evolution engaging both modalities at a highly integrated level. A temporal specialisation or advantage of one of these modalities is possible due to the development of environmental adaptations. It may be the strength of gestural hypotheses that they posit the dominance of the visual modality, at least at an early stage of protolanguage development. Further, a multimodal perspective enables the most natural integration and provides arguments in favour of the gestural approach (sections 4.2.–4.5.), by emphasising the significance of the visual modality at early stages of the human language ability.

Gesturologists such as Kendon (2011), McNeill (2012) or Sandler (2013) are firm supporters of the multimodal perspective. Their research emphasises the connection between gesture and speech (see section 1.3.1.). Gesticulations are an integral part of human conversation; therefore, we gesticulate during a phone conversation although we know that our interlocutor cannot see us (Bavelas et al. 2008). The finding that congenitally blind speakers use gesticulation provided yet another insight into the strength of gesture–speech integration (Iverson, Goldin-Meadow 1997). Neuroscientific data on the dissociation between gesticulations and signs of sign language (section 3.2.) or instrumental movements (e.g. grasping) prove the stability of gesture–speech integration. The case of I.W., who lost proprioceptive control of his body – is not able to perform instrumental movements without looking at his hands, but his gesticulations are not limited – illustrates the latter example (McNeill 2005).

The multimodal perspective has gained in popularity in various fields of

language evolution research. Primatologists point to multimodal communication patterns present in nonhuman apes. On the one hand, gestures, or communicative activities (e.g. drumming or manipulating a tree branch), are often accompanied by intentional or nonintentional production of sounds (e.g. Hobaiter, Byrne 2012). On the other hand, as Falk (2009) points out, primates' vocalisations are often linked directly to emotional state and facial expressions, which can be perceived visually. As Slocombe et al. (2011) point out, facial expressions are an important but scientifically neglected component of the "holistic information". In their literature overview, these authors have identified a tendency to adopt a simplified unimodal perspective and are now calling for others to turn to multimodal research.

Neuroscientific findings support the multimodal perspective as well – the division of vocal and visual modalities do not conform to the linguistic content of the message. Sign languages and non-linguistic communication may serve as examples. Gonseth, Vilain and Vilain (2012) argue that pointing is an activity that can be placed between modalities, where indexicality of a communicative act is achieved by means of vocal-gestural interaction. The level of the brain description seems to confirm that idea – it is not the modality that is crucial for the neural implementation of a system, but the way of processing iconicity or conventionality (Niederhut 2012). Adopting the possibility that dissociations, crucial from a language perspective, are independent from the modality is very promising to the current research in the field.

The publication is part of a project financed from the research grant No. UMO-2012/07/E/HS2/00671 of the National Science Centre.

Bibliography

Arbib, Michael A. (2002). "The mirror system, imitation, and the evolution of Language." In *Imitation in animals and artifacts*, Kerstin Dautenhahn, Chrystopher L. Nehaniv (eds.). London: A Bradford Book.

Arbib, Michael A. (2005). "From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics." *Behavioral and Brain Sciences* 28: 105–167.

Arbib, Michael A. (2012). *How the brain got language*. Oxford: Oxford University Press.

Armstrong, David F. and Sherman E. Wilcox (2007). *The gestural origin of language*. Oxford: Oxford University Press.

Armstrong, David F., William C. Stokoe and Sherman E. Wilcox (1995). *Gesture and the nature of language*. Cambridge: Cambridge University Press.

Bavelas, Janet, Jennifer Gerwing, Chantelle Sutton, Danielle Prevost (2008). "Gesturing on the telephone: Independent effects of dialogue and visibility." *Journal of Memory and Language* 58: 495–520.

Bellugi, Ursula and Edward S. Klima (2001). "Sign language." In *International encyclopedia of the social and behavioral sciences vol. 21*, Neil J. Smelser, Paul B. Baltes (eds.), 14066–14071. Oxford: Elsevier Science Publishers.

Bickerton, Derek (1990). *Language and species*. Chicago: University of Chicago Press.

Burling, Robbins (2005). *The talking ape: How language evolved*. Oxford: Oxford University Press.

Campbell, Lyle (1998). *Historical linguistics: An introduction*. Edinburgh: Edinburgh University Press.

Calvin, William H. (1982). "Did throwing stones shape hominid brain evolution?" *Ethology and Sociobiology* 3: 115–124.

Calvin, William H. (1983). *The throwing Madonna: Essays on the brain*. New York: McGraw-Hill.

Calvin, William H. and Derek Bickerton (2000). *Lingua ex machina: reconciling Darwin and Chomsky with the human brain*. Cambridge: MIT Press.

Carstairs-McCarthy, Andrew (1996). Review of Armstrong, Stokoe and Wilcox "Gesture and the nature of language." *Lingua* 99: 135–138.

Cashmore, Lisa, Natalie Uomini and Amandine Chapelain (2008). "The evolution of handedness in humans and great apes: a review and current issues." *Journal of Anthropological Sciences*, 86, s. 7–35

Corballis, Michael C. (2002). *From hand to mouth: The origins of language*. Princeton, NJ: Princeton University Press.

Corballis, Michael C. (2003). "From mouth to hand: gesture, speech, and the evolution of right-handedness." *Behavioral and Brain Sciences* 26(2): 199–208.

Corina, David P., Jyotsna Vaid and Ursula Bellugi (1992). "The linguistic basis of left hemisphere specialization." *Science* 255: 1258–1260.

Deacon, Terrence W. (1997). *The symbolic species: The coevolution of language and the brain.* New York: W.W. Norton and Company, London: Penguin Press.

Dediu Dan and Stephen C. Levinson (2013). "On the antiquity of language: the reinterpretation of Neandertal linguistic capacities and its consequences." *Frontiers in Psychology* 4: 397.

de Waal, Frans B. M. and Amy S. Pollick (2011). "Gesture as the most flexible modality of primate communication." In *The Oxford handbook of language evolution*, Maggie Tallerman, Kathleen R. Gibson (eds.), 82–89. Oxford: Oxford University Press.

Donald, Merlin (1991). *Origins of the modern mind: Three stages in the evolution of culture and cognition.* Cambridge: Harvard University Press.

Donald, Merlin (2001). *A mind so rare. The evolution of human consciousness.* New York: Norton. Emery, N.J.

Dunbar, Robin (1998). *Grooming, gossip and the evolution of language.* Cambridge: Harvard University Press.

Ekman, Paul and Wallace V. Friesen (1969). "The repertoire of nonverbal behavior: Categories, origins, usage, and coding." *Semiotica* 1: 49–98.

Emmorey, Karen (2002). *Language, cognition, and brain: Insights from sign language research.* Hillsdale, NJ: Lawrence Erlbaum.

Fadiga, Luciano, Laila Craighero and Alessandro D'Ausilio (2009) "Broca's Area in language, action and music." *Annals of the New York Academy of Sciences* 1169: 448–458.

Falk, Dean. (2009). *Finding our tongues: Mothers, infants and the origins of language.* New York: Basic Books.

Fay, Nicolas, Michael A. Arbib and Simon Garrod (2013). "How to bootstrap a human communication system." *Cognitive Science* 37: 1356–1367.

Fisiak, Jacek (1985). *Wstęp do współczesnych teorii lingwistycznych.* Warszawa: WSiP.

Fitch, W. Tecumseh (2010). *The evolution of language.* Cambridge: Cambridge University Press.

Fogassi, Leonardo and Pier Francesco Ferrari (2004). "Mirror neurons, gestures and language evolution." *Interaction Studies* 5: 345–363.

Gardner, R. Allen and Beatrice T. Gardner (1969). "Teaching sign language to a chimpanzee." *Science* 165: 664–672.

Gentilucci, Maurizio and Michael C. Corballis (2006). "From manual gesture to speech: A gradual transition." *Neuroscience and Biobehavioral Reviews* 30: 949–960.

Goldin-Meadow, Susan (2003). *Hearing gesture: How our hands helps us think*. Cambridge: The Belknap Press of Harvard University Press.

Goldin-Meadow, Susan (2005). "What language creation in the manual modality tells us about the foundations of language." *The Linguistic Review* 22: 199–225.

Goldin-Meadow, Susan (2008). "Gesture, speech, and language." In *Proceedings of the 7th International Conference on the Evolution of Language*, Andrew Smith, Kenny Smith, Ramon Ferrer-i-Cancho (eds.), 427–428. London: World Scientific.

Goldin-Meadow, Susan (2011). "What modern-day gesture can tell us about language evolution." In *The Oxford handbook of language evolution*, Maggie Tallerman, Kathleen R. Gibson (eds.), 545–557. Oxford: Oxford University Press.

Goldin-Meadow, Susan, Wing Chee So, Asli Ozy urek and Carolyn Mylander (2008). "The natural order of events: How speakers of different languages represent events nonverbally." *Proceedings of the National Academy of Sciences of the USA* 105: 9163–9168.

Gonseth, Chloe, Anne Vilain and Coriandre Vilain (2012). "Ontogeny of two communicative tools: Distance encoding and multimodality in deictic pointing." In *The Evolution of Language. Proceedings of the 9th International Conference (EVOLANG 9)*, Thomas C. Scott-Phillips, Mónica Tamariz, Erica A. Cartmill, James R. Hurford (eds.), 150–157. New Jersey: World Scientific.

Harrison, Rebecca (2008). "What can the study of handedness in nonhuman apes tell us about the evolution of language." In *Proceedings of the 7th International Conference on the Evolution of Language*, Andrew Smith, Kenny Smith, Ramon Ferrer-i-Cancho (eds.), 431–432. World Scientific.

Hauser, Marc D., Noam Chomsky and W. Tecumseh Fitch (2002). "The faculty of language: What is it, who has it, and how did it evolve?" *Science* 298 1569–1579.

Hewes, Gordon W. (1973). "Primate communication and the gestural origins of language." *Current Anthropology* 14: 5–24.

Hewes, Gordon W. (1981). "Pointing and language." In *The cognitive representation of speech*, Terry Myers, John Laver, John Anderson (eds.), 263–269. Amsterdam: North-Holland.

Hewes, Gordon W. (1996). "A history of the study of language origins and the gestural primacy hypothesis." In *Handbook of human symbolic evolution*, Andrew Lock, Charles R. Peters (eds.), 571–595. Oxford: Oxford University Press.

Hickok, Gregory, Ursula Bellugi and Edward S. Klima (1996). "The neurobiology of sign language and its implications for the neural basis of language." *Nature* 381: 699–702.

Hobaiter, Catherine and Richard W. Byrne (2012). "Gesture use in consortship." *Developments in Primate Gesture Research* 6: 129–146.

Hockett, Charles F. (1960). "The origin of speech." *Scientific American* 203: 88–96.

Hopkins, William D. (2006). "Comparative and familial analysis of handedness in great apes." *Psychological Bulletin* 132: 538.

Hurford, James R. (2007). *The origins of meaning. Language in the light of evolution*. Oxford: Oxford University Press.

Hutto, Daniel D. (2008). "First communions: Mimetic sharing without theory of mind." In *The Shared Mind: Perspectives on intersubjectivity*, Jordan Zlatev, Timothy P. Racine, Chris Sinha, Esa Itkonen (eds.), 246–276. Amsterdam: John Benjamins Publishing.

Iacoboni, Marco, Roger P. Woods, Marcel Brass, Harold Bekkering, John C. Mazziotta and Giacomo Rizzolatti (1999). "Cortical mechanisms of human imitation." *Science* 286: 2526–2528.

Iverson, Jana M. and Susan Goldin-Meadow (1997). "What communication got to do with it? Gesture in congenitally blind children." *Developmental psychology* 33: 453–467.

Johansson, Sverker (2005). *Origins of language. Constraints on hypothesis*. Amsterdam: John Benjamins Publishing.

Johansson, Sverker (2012). "The case for Neanderthal language – how strong is it?" In *The Evolution of Language. Proceedings of the 9th International Conference (EVOLANG 9)*, Thomas C. Scott-Phillips, Mónica Tamariz, Erica A. Cartmill, James R. Hurford (eds.), 173–180. New Jersey: World Scientific.

Kendon, Adam (1991). "Some considerations for a theory of language origins." *Man* 26: 199–221.

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

Kendon, Adam (2008). "Signs for Language Origins?" *The Public Journal of Semiotics* II(2): 2–29.

Kendon, Adam (2011). "Some modern considerations for thinking about language evolution: A discussion of The Evolution of Language by Tecumseh Fitch." *The Public Journal of Semiotics* III(1): 79–108.

Kenneally, Christine (2007). *The first word*. New York: Viking.

Kegl, Judy, Ann Senghas and Marie Coppola (1999). "Creation through contact: Sign language emergence and sign language change in Nicaragua." In *Language Creation and Language Change: Creolization, Diachrony, and Development*, Michel DeGraff (ed.), 179–237. Cambridge: MIT Press.

Knecht, Stefan, Bianca Dr ager, Michael Deppe, Lars Bobe, Hubertus Lohmann, Agnes Fl oel, Bernd E. Ringelstein and Henning Henningsen (2000). "Handedness and hemispheric language dominance in healthy humans." *Brain* 123: 2512–2518.

Knight, Chris (2000). "Play as precursors of phonology and syntax." In *The evolutionary emergence of language*, Chris Knight, Michael Studdert-Kennedy, James R. Hurford (eds.), 99–119. Cambridge: Cambridge University Press.

Krauss, Robert M., Yihsiu Chen and Rebecca F. Gottesman (2000). "Lexical gestures and lexical access: A process model." In David McNeill (ed.), *Language and gesture* (pp. 261–283). New York: Cambridge University Press.

Liberman, Alvin Meyer, Franklin S. Cooper, Donald P. Shankweiler and Michael Studdert-Kennedy (1967). "Perception of the speech code." *Psychological review* 74: 431–461.

Liberman, Alvin Meyer, Ignatius G. Mattingly (1985). "The motor theory of speech perception revisited." *Cognition* 21: 1–36.

Liberman, Alvin Meyer, Doug H. Whalen (2000). "On the relation of speech to language." *Trends in cognitive sciences* 4: 187–196.

Lieberman, Philip (2003). "Motor control, speech, and the evolution of human language." In *Language evolution*, Morten H. Christiansen, Simon Kirby (eds.), 255–271. New York: Oxford University Press.

MacNeilage, Peter F. (1998). "The frame/content theory of the evolution of speech production." *Behavioral and Brain Sciences* 21: 499–556.

MacNeilage, Peter F. (2008). *The origin of speech*. Oxford: Oxford University Press.

Maestripieri, Dario (2007). "Gestural communication in three species of macaques (Macaca mulatta, M. nemestrina, M. arctoides): Use of signals in relation to dominance and social context." In *Gestural communication in nonhuman and human primates*, Katja Liebal, Cornelia M uller, Simone Pika (eds.), 53–68. Amsterdam: John Benjamins Publishing.

McGurk, Harry and John MacDonald (1976). "Hearing lips and seeing voices." *Nature* 264: 746–748.

McNeill, David (1992). *What gestures reveal about thought*. Chicago: The University of Chicago Press.

McNeill, David (2005). *Gesture and thought*. Chicago: The University of Chicago Press.

McNeill, David (2012). *How language began: Gesture and speech in human evolution*. Cambridge: Cambridge University Press.

Meguerditchian, Adrien, Hélène Cochet and Jacques Vauclair (2011). "From gesture to language: ontogenetic and phylogenetic perspectives on gestural communication and its cerebral lateralization." In *Primate communication and human language: vocalisation, gestures, imitation and deixis in humans and non-humans*, Anne Vilain, Jean-Luc Schwartz, Christian Abry, Jacques Vauclair (eds.), 89–118. Amsterdam: John Benjamins Publishing.

Mithen, Steven (2005). *The singing Neanderthals: The origins of music, language, mind and body*. London: Weidenfeld and Nicholson.

Murata, Akira, Luciano Fadiga, Leonardo Fogassi, Vittorio Gallese, Vassilis Raos and Giacomo Rizzolatti (1997). "Object representation in the ventral premotor cortex (area F5) of the monkey." *Journal of Neurophysiology* 78(4): 2226–30.

Niederhut, Dillon (2012). "Gesture and the origin of language." *The Evolution of Language. Proceedings of the 9th International Conference (EVOLANG 9)*, Thomas C. Scott-Phillips, Mónica Tamariz, Erica A. Cartmill, James R. Hurford (eds.), 266–273. New Jersey: World Scientific.

Orzechowski, Sylwester, Sławomir Wacewicz and Przemysław Żywiczyński (2016). "Vocal-auditory feedback and the modality transition problem in language evolution." *Reti, saperi, linguaggi. Italian Journal of Cognitive Sciences* 3(1): 157–178.

Petitto, Laura Ann and Paula F. Marentette (1991). "Babbling in the manual mode: Evidence for the ontogeny of language." *Science* 251: 1493–1496.

Pika, Simone (2008). "What is the nature of the gestural communication of great apes?" In *The Shared Mind: Perspectives on intersubjectivity*, Jordan Zlatev, Timothy P. Racine, Chris Sinha, Esa Itkonen (eds.), 165–186. Amsterdam: John Benjamins Publishing.

Pika, Simone and Katja Liebal (2006). "Differences and similarities between the natural gestural communication of the great apes and human children." In *Proceedings of the 6th International Conference on the Evolution of Language (EVOLANG 6)*, Angelo Cangelosi, Andrew D. M. Smith, Kenny Smith, 267–274. World Scientific Publishing Company.

Pika, Simone, John C. Mitani (2009). "The directed scratch: Evidence for a referential gesture in chimpanzees?" In *The prehistory of language*, Rudolf Botha, Chris Knight (eds.), 166–180. Oxford: Oxford University Press.

Pika, Simone, Katja Liebal, Josep Call and Michael Tomasello (2005). "The gestural communication of apes." *Gesture* 5: 41–56.

Pinker, Steven and Ray Jackendoff (2005). "The faculty of language: what is special about it?" *Cognition* 95: 201–236.

Pollick, Amy S. and Frans B. M. de Waal (2007). "Ape gestures and language evolution." *Proceedings of the National Academy of Sciences of the USA* 104: 8184–8189.

Rizzolatti, Giacomo and Michael A. Arbib (1998). "Language within our grasp." *Trends in Neurosciences* 21: 188–194.

Rizzolatti, Giacomo, Luciano Fadiga, Leonardo Fogassi and Vittorio Gallese (1996). "Premotor cortex and the recognition of motor actions." *Cognitive Brain Research* 3: 131–141.

Rowe, Meredith L. and Susan Goldin-Meadow (2009). "Early gesture selectively predicts later language learning." *Developmental Science* 12: 182–187.

Sandler, Wendy (2013). "Vive la différence: Sign language and spoken language in language evolution." *Language and Cognition* 5(2–3): 189–203.

Slocombe, Katie (2011). "Have we underestimated great ape vocal capacities?" In *The Oxford Handbook of Language Evolution*, Maggie Tallerman, Kathleen R. Gibson (eds.), 90–95. Oxford: Oxford University Press.

Slocombe, Katie, Bridget Waller and Katja Liebal (2011). "The language void: the need for multimodality in primate communication research." *Animal Behaviour* 81(5): 919–924.

Stokoe, William C. (1960). *Sign language structure*. Silver Spring: Linstok Press.

Stokoe, William C. (1991). "Semantic phonology." *Sign Language Studies* 71: 107–114.

Stokoe, William C. (2001). *Language in hand: Why sign came before speech*. Washington, DC: Gallaudet University Press.

Stokoe, William C., Dorothy C. Casterline and Carl G. Croneberg (1965). *A dictionary of American Sign Language on linguistic principles*. Silver Spring: Linstok.

Studdert-Kennedy, Michael (2005). "How did language go discrete?" In *Language origins – perspectives on evolution*, Maggie Tallerman (ed.), 48–67. Oxford: Oxford University Press.

Summerfield, A. Quentin (1992). "Lipreading and audio-visual speech perception." *Philosophical Transactions of the royal society B* 335: 71–78.

Sutton-Spence, Rachel (2005). *Analysing Sign Language Poetry*. New York: Palgrave Macmillan.

Tallerman, Maggie (2011). "Protolanguage." In *The Oxford Handbook of Language Evolution*, Maggie Tallerman, Kathleen R. Gibson (eds.), 479–491. Oxford: Oxford University Press.

Tallerman, Maggie and Kathleen R. Gibson (eds.). (2011). *The Oxford handbook of language evolution*. Oxford: Oxford University Press.

Tinbergen, Nikolaas (1951). *The study of instinct*. New York: Oxford University Press.

Tomasello, Michael (2008). *Origins of human communication*. Cambridge, MA: MIT Press.

Wacewicz, Sławomir (2008). "Ewolucja języka: horyzont metodologiczny." In *Metodologie językoznawstwa. Współczesne tendencje i kontrowersje*, Piotr Stalmaszczyk (ed.), 27–42. Kraków: Wydawnictwo LEXIS.

Wacewicz, Sławomir (2013). "Ewolucja języka – współczesne kontrowersje." In *Metodologie językoznawstwa. 1. Ewolucja języka. Ewolucja teorii językoznawczych*, Piotr Stalmaszczyk (ed.), 11–26. Łódź: Wydawnictwo Uniwersytetu Łódzkiego.

Wacewicz, Sławomir and Przemysław Żywiczyński (2008). "Broadcast transmission, signal secrecy and gestural primacy hypothesis." In *The Evolution of*

Language. Proceedings of the 7th International Conference (EVOLANG 7), Andrew Smith, Kenny Smith, Ramon Ferrer-i-Cancho (eds.), 354–361. Singapore: World Scientific.

Wacewicz, Sławomir, Przemysław Żywiczyński and Sylwester Orzechowski (2016). "Visible movements of the orofacial area." *Gesture* 15(2): 250–282.

Zlatev, Jordan (2008). "The co-evolution of intersubjectivity and bodily mimesis." In *The Shared Mind: Perspectives on intersubjectivity*, Jordan Zlatev, Timothy P. Racine, Chris Sinha, Esa Itkonen (eds.), 215–244. Amsterdam: John Benjamins Publishing.

Zlatev, Jordan (in press). *Bodily mimesis and the transition to speech.*

Żywiczyński, Przemysław and Sławomir Wacewicz (2008). *Ewolucja języka. W stronę hipotez gesturalnych.* Toruń: Wydawnictwo Naukowego Uniwersytetu Mikołaja Kopernika.

Ethnologue: `http://www.ethnologue.com/`

World Atlas of Linguistic Structures: `http://wals.info/`

Originally published as "Problem 'zmiany modalności' w hipotezach pierwszeństwa gestów w ewolucji języka: w stronę hipotez multimodalnych," Studia Semiotyczne 28–29 (2015), 335–369. Translated by Monika Boruta.