Situating Language in the Real-World: The Role of Multimodal Iconicity and Indexicality

MARGHERITA MURGIANO

YASAMIN MOTAMEDI

GABRIELLA VIGLIOCCO [iD]

*Author affiliations can be found in the back matter of this article

ABSTRACT

In the last decade, a growing body of work has convincingly demonstrated that languages embed a certain degree of non-arbitrariness (mostly in the form of iconicity, namely the presence of imagistic links between linguistic form and meaning). Most of this previous work has been limited to assessing the degree (and role) of non-arbitrariness in the speech (for spoken languages) or manual components of signs (for sign languages). When approached in this way, non-arbitrariness is acknowledged but still considered to have little presence and purpose, showing a diachronic movement towards more arbitrary forms. However, this perspective is limited as it does not take into account the situated nature of language use in face-to-face interactions, where language comprises categorical components of speech and signs, but also multimodal cues such as prosody, gestures, eye gaze etc. We review work concerning the role of context-dependent iconic and indexical cues in language acquisition and processing to demonstrate the pervasiveness of non-arbitrary multimodal cues in language use and we discuss their function. We then move to argue that the online omnipresence of multimodal non-arbitrary cues supports children and adults in dynamically developing situational models.

1. INTRODUCTION

Moving away from a more traditional view of language (e.g., de Saussure, 1916; Hockett, 1960), numerous studies in recent years have focused on non-arbitrary features of language such as iconicity (i.e., the presence of imagistic links between some features of the linguistic form and attributes of the corresponding referent) in both spoken and sign languages (see for example Dingemanse et al., 2015; Perniss et al., 2010 for reviews), and the indexicality (i.e., signs that index, point to, specific referents) of certain signs like pronouns in sign languages (e.g., the sign for the pronoun "you" in many sign languages consists in an extended index finger just like in pointing toward the person) (Liddell, 2003; Johnston, 2013; Meier & Lillo-Martin, 2013).

However, human languages are still thought to be primarily arbitrary, with properties like iconicity and indexicality considered as marginal (Dingemanse, 2018). Non-arbitrariness is still generally argued to be a negligible feature, uncommon in the majority of studied linguistic systems — particularly English. Even in sign languages, where non-arbitrariness is indisputably more visible, it has been hypothesised to reduce over time in favour of more arbitrary forms (Emmorey, 2014; Frishberg, 1975; Klima & Bellugi, 1979) or in any case not to impact language learning or use (e.g., Pettito, 1987). This paper offers a different perspective on the topic, a perspective that we believe has far reaching consequences not only for how we characterise the presence, function and limits of non-arbitrariness in language, but language itself.

Previous work investigating non-arbitrariness has invariably focussed on language **as a population-level system** in which the structured, categorical components of speech (for spoken languages) and manual signs (for sign languages) can be described in a population at a given time point, to characterise the language synchronically and track diachronic change. In this way, language has been empirically studied as speech/sign or text, separately from other contextual components of communication. While this perspective has been undeniably fruitful, we believe it can only lead to a partial understanding of language and its features because it offers a distilled abstraction of how language really manifests.

Linguistic use is first of all inevitably embedded in a *physical context*, the environment in the here and now which can disambiguate the message or offer useful resources to enhance its communicative power (Clark, 2016; Mondada, 2019). Most importantly, while language certainly exists as the categorical components of speech and signs, these components are never alone in the situated *communicative context* where language is learnt and used: online, face-to-face interactions present a composite and richer message exploiting simultaneously multiple articulators and channels. Other cues, such as gesture, prosody and eye gaze (to name a few) are always present, have been argued to be essential for the phylogenetic and ontogenetic emergence of language, and are key components of meaning-making (Holler & Levinson, 2019; Kendon, 2012, 2014; Perniss, 2018; Vigliocco et al., 2014).

In this paper, along with other scholars (e.g., Hasson et al., 2018; Holler & Levinson, 2019), we argue that we should broaden our lens beyond a notion of language as something that can be investigated in a single communicative channel (e.g., vocal *or* manual) to the study of **language as situated**, namely, as the ensemble of speech (or sign) in specific communicative and physical contexts as dynamically presenting during communicative interactions. This perspective forces a rethinking of the traditional distinction between what we consider as linguistic and non-linguistic (Fontana, 2008; Kendon, 2012, 2014; Slobin, 2008). Similarly, it pushes us to rethink what we consider a 'core' feature of language and what is instead a secondary or even negligible attribute. Thus, when looking at *language as a system*, non-arbitrariness in the linguistic form appears to be a marginal feature that decreases over time as the result of pressures to (for example) reduce production effort and memory demands.

However, when looking at *language as situated*, the presence of non-arbitrariness is much more than marginal: during face-to-face interactions language users draw from both linguistic resources available in the system and other online multimodal resources such as iconic gestures and iconic prosody (Drijvers & Özyürek, 2017; Herold et al., 2012; Iverson et al., 1999; Vigliocco et al., 2019) as well as points and eye gaze (Cooperrider, 2016; Holler et al., 2014; Motamedi et al., 2020). Taken together, these iconic and indexical multimodal cues provide effective mechanisms to single out and bring "to the mind's eye" referents being talked about.

Note here that we distinguish iconicity and indexicality as non-arbitrary components that have been undervalued under the language as a system approach. Another component of human language that is often included along with iconicity and indexicality when discussing non-arbitrariness is systematicity, which refers to regular correspondences between form and meaning, without the form having to represent the meaning through resemblance or analogy (Dingemanse et al., 2015; Monaghan et al., 2014). However, we suggest that systematicity is qualitatively different from both iconicity and indexicality. Where iconicity and indexicality can exist independently from a system (i.e., a spontaneous pointing gesture can be understood without reference to a pointing system), systematic correspondences between form and meaning can only be understood in relation to the whole system. As such, systematicity is best understood primarily under the language as a system view, and we therefore focus on iconicity and indexicality as revealed through the language as situated perspective.

Below, we first briefly review research on non-arbitrariness carried out in the language as a system tradition, spelling out the main shortcomings of such an approach. We then introduce the perspective of language as situated by outlining how language, as it is used in face-to-face, interactive settings, is dynamic, multimodal and contextualised. We discuss the implications of this approach for our understanding of iconicity and indexicality in two research domains— language acquisition and language processing— as examples of areas where considering language only at the system level is especially problematic because both are for the most part carried out in face-to-face contexts where all the multimodal cues are available. In both cases, considering language as only speech or only sign leaves us with an impoverished view of how humans learn and process language.

2. NON-ARBITRARINESS FROM THE LANGUAGE AS A SYSTEM PERSPECTIVE

Research on iconicity has seen a boom over the last decade, with an increasing acknowledgement of iconicity as a non-trivial property of language (Perniss et al., 2010), present in both signed and spoken languages (Dingemanse, 2018; Dingemanse et al., 2015; Perniss et al., 2010). As already mentioned, most investigations have focussed on languages as structural, rule-based systems of form-meaning mappings, reflecting population-level biases for ease of articulation, ease of learning and communicative efficiency. Thus, research first has established the presence of iconicity at different levels of description. At the lexical level, whole words or signs are judged to be iconic (i.e., 'moo' sounds like the noise a cow makes, or the British Sign Language (BSL) sign for BOOK represents the leaves of a book; Boyes-Braem, 1986; Perry et al., 2015; Pietrandrea, 2002; Vinson et al., 2008; Winter et al., 2017). At the sub-lexical level, meaningful correspondences exist between particular phonemes or certain acoustic properties and particular semantic properties, such as the vowel *i:* being associated with small size in spoken languages (Knoeferle et al., 2017; Marchand, 1959). Such correspondences also exist in sign languages, with phonological parameters such as movement, location or handshape being iconically linked to the meaning of the sign (Boyes-Braem, 1981; Brentari, 2012; Emmorey & Herzig, 2003; Thompson et al., 2010). Finally, iconicity has been studied at the syntactic level, where, for example, the structure of the signed or spoken phrase can represent the structure of the event in both spoken and sign languages (Christensen et al., 2016; Diessel, 2008; Strickland et al., 2015; Wilbur, 2004).

In addition to having confirmed that iconicity is present across languages, research has established that iconicity plays a role in language learning and processing. For instance, iconicity has been shown to be common in children's early vocabularies (Caselli & Pyers, 2019; Perry et al., 2017; Tardif et al., 2008; D. P. Vinson et al., 2008) and argued to facilitate word and sentence-level comprehension in toddlers and young children (De Ruiter et al., 2018; Imai et al., 2008). In adult language users, studies have shown that adults presented with sound-symbolic words from unfamiliar languages (e.g., Japanese, Semai) can guess their meanings above what would be expected by chance (Dingemanse et al., 2016; Lockwood et al., 2016). Though the role of iconicity in lexical acquisition is still the subject of a lively debate (see for example Ortega, 2017 for a review), most recent research has accumulated evidence suggesting that iconicity helps to ground referential communication providing information about properties of

real-world referents that may be of particular help for young language learners (Imai & Kita, 2014) and for users and learners of emerging linguistic systems (Fay et al., 2013; Perlman et al., 2015; Roberts et al., 2015).

Indexicality, in contrast, is largely ignored by the language as a system approach. At the lexical level, demonstrative pronouns have been considered indexical, such that they index a meaning without using its conventional lexical form. However, this differs somewhat from our definition of indexical as providing a visual link to the intended referent (e.g., through a finger point). Pointing, and other forms of deixis have largely been ignored from a systemic point of view for spoken languages. In sign language research, indexical points have long been understood as part of the grammatical systems of most sign languages. As a consequence, they are often studied under a language as a system view, that posits grammatical pointing as qualitatively different from gestural points. As such, the overlap between gestural and grammatical points has been argued not to be understandable by children. Pettito (1987) argues that the finding that children learning ASL can erroneously produce "you" (point toward you) when "I" (point toward me) is intended, just like children speaking English, provides evidence for the independence of grammatical and gestural systems. More recently, that perspective has shifted, to understand the similarities between grammatical and gestural pointing (Cormier et al., 2013), where grammatical points, though very similar to gestural points, are systematically constrained by other aspects of the linguistic system (Fenlon et al., 2019).

Understanding the effects of non-arbitrariness on language learning also highlights what non-arbitrariness *cannot* do. For example, from a system-wide perspective, iconicity becomes limiting when there is an asymmetry between the dimensions of the meaning space and the dimensions of the signal space (Gasser, 2004; Little et al., 2017): i.e., as a language grows in the number of meanings it wants to refer to, it becomes more difficult to maintain iconicity across the system. Furthermore, iconicity purportedly hinders discriminability in crowded meaning spaces, such that lexical items that occur in dense semantic networks tend to be less iconic (Sidhu & Pexman, 2018), and the specified nature of iconic forms, referring to particular properties of referents, may mean that iconic forms are less well-suited to refer to more general or more abstract concepts (Lupyan & Winter, 2018). These observations suggest that from the systemic perspective, iconicity may serve little purpose beyond first language acquisition, and beyond the initial grounding and early evolution of novel referential systems. That is, in mature languages and speakers, the remaining iconic forms are 'relics' of our (phylogenetic and ontogenetic) history. This hypothesis has some support from natural language data. Studies of children's early vocabulary and from child-directed language, based on English-speaking populations, suggest that the proportion of iconic words reduce in both cases as children get older (Laing, 2014; Perry et al., 2017; Vigliocco et al., 2019). For understanding the evolution of iconic forms, Frishberg's (1975) study of signs from American Sign Language (ASL) documented a movement from more iconic to more abstract forms, a pattern supported by results from experimental studies of novel communication systems (Garrod et al., 2007; Theisen et al., 2010).

Yet, there are at least some studies that have shown processing effects of iconicity in adult language users. First, there is evidence that adult speakers can identify correspondences between sounds and meanings (e.g., Köhler, 1929; Ramachandran & Hubbard, 2001). For example it has been shown that Dutch and English speakers can guess the meaning of sound-symbolic Japanese words above chance (Lockwood et al., 2016; Oda, 2000) and English speakers have been shown to map Japanese sound-related ideophones to similar meanings as Japanese native speakers (Iwasaki et al., 2007). Lexical iconicity has also been shown to facilitate processing in lexical decision tasks; for example, Sidhu et al. (2020) found that iconic words were processed more quickly in a visual lexical decision tasks by healthy English-speaking adults, and Meteyard et al. (2015) found a facilitation effect in patients with aphasia in an auditory lexical decision task, with participants recognising iconic words faster than non-iconic ones. In sign language, Vinson et al (2015) showed that iconic signs in BSL are produced faster than less-iconic ones. These, however, are rather limited effects.

As already introduced, there are several important shortcomings to the system-level view of language that may be critical for our discussion. First, work in this tradition is shaped by the languages from western-industrialised communities (especially English) on which it is based, and which have more limited sound-symbolic and iconic vocabularies than other linguistic

systems, such as non-Indo-European languages and sign languages (Dingemanse, 2018; Vigliocco et al., 2014). As such, even within the system-level view, the prevalence of non-arbitrariness across the world's languages may have been underestimated. Second, this view fails to account for the very systematic and dynamic nature of language behaviours it aims to capture. By focusing on the static properties of language (e.g., either lexical *or* phonological *or* syntactic iconicity), we do not account for the multiplex ways in which behaviours associated with language use interact. For example, evidence showing that the proportion of iconic words in children's vocabularies declines with age may support the view that iconicity is less useful beyond early word learning, but little to no data exists that tracks the use of non-arbitrary forms in other modalities (such as iconic and pointing gestures), or how children might draw from information in the lexical and gestural channels simultaneously.

Consequently, we propose that a more ecological model of language use, one that accounts for face-to-face interaction and multimodal behaviour, is imperative to gain a more comprehensive understanding of the role that non-arbitrariness plays in both language learning and processing. Furthermore, given that the human language capacity likely evolved from pre-linguistic communicative interactions (Levinson & Holler, 2014), an understanding of the contextual constraints of language will further shed light on the ways in which the structures and components of language are adapted for learning and use in interaction (Chater & Christianesen, 2008; Kirby et al., 2015).

Importantly, the narrow focus of the language as a system approach is not only an issue for research on non-arbitrariness, but cognitive science in general. Indeed, there is increasing call among language scientists for more ecologically valid and multiplex approaches, such as those concerning language processing (Holler & Levinson, 2019), neurobiology (Hasson et al., 2018) and language development (Rączaszek-Leonardi et al., 2018). In the following sections, we use non-arbitrariness to highlight how a focus on language as a stable and static rule-based system fails to account for the rich set of non-arbitrary behaviours that govern natural language learning and use, and suggest that understanding the contextual constraints and affordances of face-to-face interaction can offer a more comprehensive picture.

3. NON-ARBITRARINESS FROM THE LANGUAGE AS SITUATED PERSPECTIVE

Within linguistics, pragmatics has investigated language use in context, showing that the access to a message doesn't solely rely on the knowledge of underlying lexical and grammatical rules, but on a larger range of situational factors. Among these factors shown to affect language comprehension are the cultural assumptions and habits shaping a shared 'common ground' between speaker and listener (Bohn & Köymen, 2018; Clark, 1996) and socio-pragmatic skills, e.g., the ability to infer communicative intentionality (Grice, 1975). This approach has revealed language as a joint practice with which interlocutors cooperatively signify reality (Clark, 1996; Tomasello, 2008) and act on it, for example influencing others' behavior – like asking a colleague to close the window just by saying "it's very cold in the office today!" – or performing proper acts, like when we promise, demand or forbid something just by saying so (Austin, 1962; Searle, 1969).

However, shedding light on the contextual nature of linguistic comprehension also means acknowledging that language use is, for the most part, embedded in face-to-face interactions where language manifests through the combination of speech and signs and other simultaneous embodied resources that allow us to convey meanings. Levinson and Holler (2014) argue that human language is the result of an evolutionary process originated in situated communicative interactions that have led it to evolve as a 'system of systems' where various expressive channels developed phylogenetically to contribute with their different strengths to the communicative goal. We observe this multi-layered nature of language everyday, engaging simultaneously the audio-vocal, the visual-gestural and the oro-facial channel, and with listeners and learners in both spoken and sign languages extracting information from a multimodal communicative context. In spoken languages, the linear speech component is combined with concurrent mouth movements that influence speech perception (Chandrasekaran et al., 2009; Mcgurk & Macdonald, 1976) and prosodic modulations which mark information in beneficial ways (Cutler et al., 1986, 1997; Shintel et al., 2014). Gestures naturally co-occurring with speech have been largely recognized as an integral part of the message (Abner et al., 2015; Kendon, 2004; McNeill, 1992), along with other cues such as eye gaze (Staudte & Crocker, 2011), as well as facial

expressions and body movements. Brow movements, facial expressions and posture shifts have grammatical and lexical functions in the majority of sign languages (Mohr, 2014; Sandler, 1999), and are in some cases necessary to distinguish between meanings (Pfau & Quer, 2010; Woll, 2014). Finally, when using language we also frequently resort to the surrounding physical context, for example manipulating objects to demonstrate how they work or using them as props to represent something else (Clark, 2016; Mondada, 2019). In this way, object-directed actions are frequently co-opted in communication to offer visual information about the message (Alibali et al., 2011; Brand et al., 2002, 2007, S. Kelly et al., 2015; Vigliocco, Motamedi, et al., 2019).

In short, language used in face-to-face online communication is a multimodal phenomenon (Kendon, 2014; Perniss, 2018; Vigliocco et al., 2014) enacted through the combination of different resources – speech/signs, body gestures and object manipulations – the use of which is pervasive and, we argue, advantageous to comprehension and learning. In the language as situated framework, these features are not defined negatively (e.g., "non-linguistic" signals, "non-manual" components), but are instead conceived as part and parcel of language (Kendon, 2012, 2014; Liddell, 2003; Perniss, 2018; Slobin, 2008; Vigliocco et al., 2014). For this reason, the language as a system view is not incompatible with the perspective proposed here: rather, the latter includes the former, with language as a structure of categorical components being part of a broader, diversified ensemble that constitutes language use situated in the communicative and physical context.

If we take the language as situated view, we are necessarily pushed to questioning the dogma of language as primarily arbitrary. The multimodal components pervasively accompanying linguistic exchanges often provide non-arbitrary relations to the meaning in the form of both iconicity — exhibiting the qualitative features of the referent in the communicative form — and indexicality — creating an associative visual link with the referent (Peirce, 1974). Iconic mappings, while available — as we pointed out — in the linguistic repertoire at different levels, can be also exploited *in-situ* (Holler & Levinson, 2019). For example, spoken languages exploit iconic gestures performed with the hands (Kendon, 2008; McNeill, 1992) or the whole body (Clark, 2016; Stec et al., 2016), as well as prosodic modulations such as slowing the speech rate to refer to a slow action (Nygaard et al., 2009). Sign languages exhibit similar prosodic modification of signs; e.g., slower motion to indicate an effortful action (Perniss, 2018) as well as channel-specific phenomena such as role-taking; e.g., a narrator shifting to the viewpoint of the actor (Cormier et al., 2015). Furthermore, eye gaze movements following the position of an object, object manipulations or pointing gestures can be used to index referents in both sign and spoken languages. We illustrate the range of cues we discuss here under the language as situated framework in *Figure 1*.
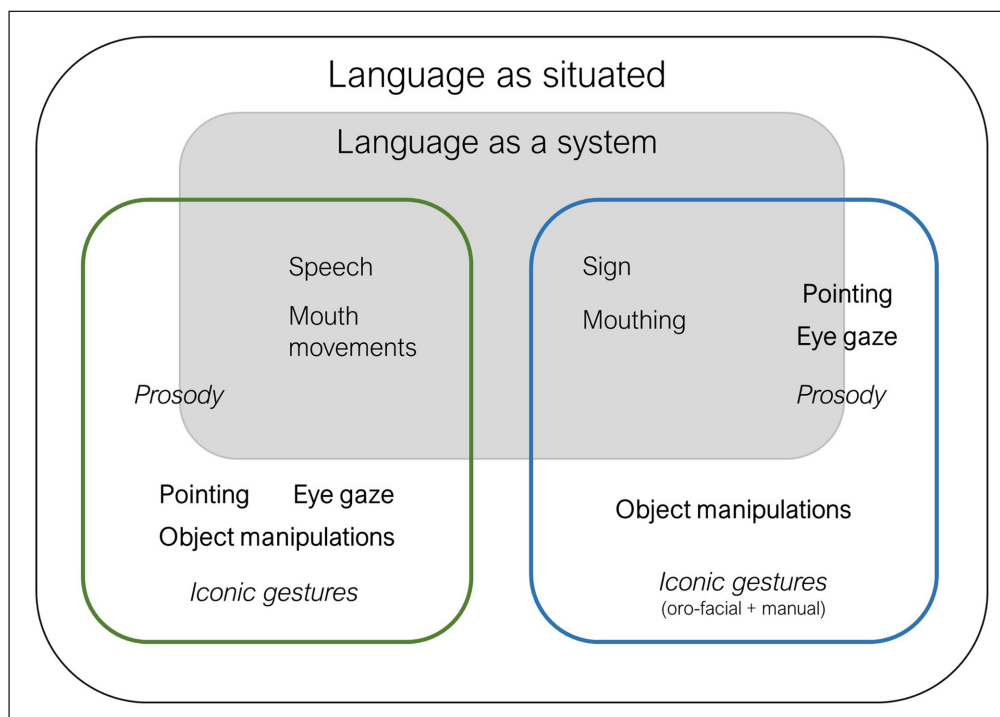

Figure 1 Components of language under a language as situated perspective. Language as a system remains in place under this view, containing those behaviours characterised as systemic. We show different communicative behaviours for both signed (blue square) and spoken (green square) languages. Iconic cues are shown in italics, indexical cues in bold. Some features (e.g., pointing in sign languages) can be characterised as either systemic or contextual.

We suggest that looking at language as situated provides a more ecologically valid framework necessary to fully understand how the different components of language are employed during face-to-face interactions in the online physical and communicative context. This view reveals human communication as a rich set of arbitrary and non-arbitrary components used in context-sensitive ways to fulfill different communicative goals. In particular, it underscores non-arbitrary communicative mechanisms in the form of indexical and iconic multimodal cues as pervasively exploited by language users to reinforce the denotative function of the message and enhance its depictive power (Clark, 1996, 2016; Sallandre & Cuxac, 2001), allowing interlocutors to converge on a situational model. Below, we review how caregivers and children exploit non-arbitrary multimodal cues to fulfill these functions in situated communication in ways that support language acquisition.

4. NON-ARBITRARINESS IN LANGUAGE ACQUISITION

Over the years, research has highlighted the dialogical and embodied nature of language development, stressing the fundamental role played by contextual cues (Bohn & Frank, 2019; Rączaszek-Leonardi et al., 2018). For example, a study by Cartmill et al. (2013) illustrated that referential transparency (operationalized as how clearly the meaning of words can be inferred from accompanying contextual cues) is an important predictor of vocabulary size at a later age (54 months). The physical context in which children learn provides critical cues, such as the objects being referred to and the affordances for actions being talked about. However, the communicative context also provides a wealth of useful cues, many of which afford iconic and indexical strategies.

Caregivers use iconicity in interactions with their children: they consistently use prosodic modulations to facilitate the interpretation of contrasting meanings (Herold, Nygaard, & Namy, 2011). Furthermore, when asked to produce novel adjectives, adults using child-directed language rely on systematic sound-to-meaning correspondences (Nygaard et al., 2009) that preschool children successfully use to infer meaning when no other cues are present (Herold, Nygaard, Chicos, et al., 2011). Manual iconicity has also been found to support learning: hearing two-and-a-half-year olds have been shown to be able to recognize iconic mappings embedded in unfamiliar signs representing actions (Tolar et al., 2008) and to learn iconic gestures at 26 months, when they exhibit difficulties in learning arbitrary gestural labels (Namy et al., 2004). Furthermore, iconic co-speech gestures have been shown to facilitate verb learning in children as young as 2 (Goodrich & Hudson Kam, 2009) and to influence, with their depictive features, the generalizations of ambiguous novel labels in 3-year olds (Mumford & Kita, 2014).

Caregivers also use indexical cues. Caregiver eye gaze has been shown to work as an attention-getter (Senju & Csibra, 2008) that children seem to be able to use early on in development (D'Entremont, 2000; Hofsten et al., 2005; Senju & Csibra, 2008). Points are the most common gesture used by caregivers very early on, and it has been found that their production in association with the label correlates with children's vocabulary development (Iverson et al., 1999; O'Neill et al., 2005; Rowe & Goldin-Meadow, 2009). Object manipulations have also been considered to provide a useful cue: caregivers have been found to exaggerate their hand actions when focusing on new objects, enhancing features like repetitiveness and range of motion in ways thought to facilitate children's attention and learning (Brand et al., 2002, 2007).

Considering iconicity and indexicality from the language as situated perspective allows us to go beyond asking whether non-arbitrariness in caregivers' input may or not support learning to asking questions concerning under which conditions non-arbitrariness will be most useful to children. Most of the work described above has considered lexical acquisition in situations in which referents are physically present and the child has to identify the correct referent in ambiguous contexts (Quine, 1960; Snedeker & Gleitman, 2004). However, referents are not always present: caregivers can – and often do (Tomasello & Kruger, 1992) – talk about spatially and temporally displaced referents (e.g., the toy in the other room or the walk in the park that is about to happen), and these displaced scenarios can also provide learning opportunities. Experimental studies indicate that children learn from displaced contexts better (Tomasello & Kruger, 1992) or at least equally well than in joint attentional contexts (Tomasello et al., 1996; Tomasello & Barton, 1994), but little is known about *how* children can learn in displaced contexts. We propose that a close analysis of non-arbitrary multimodal strategies used by caregivers in face-to-face communication may provide us with important insight.

Perniss et al. (2017) asked whether caregivers modify iconic BSL signs more often when referents were displaced than when they were visually available. Caregivers were presented with sets of toys in one condition, and asked to imagine (the child was not present) talking to their child about the toys. In a second condition, the same caregivers had to imagine talking to their child about the toys, but without the toys present. They found that while caregivers used points more often to refer to the toys when they were present, they modified iconic signs more often (by enlargement, lengthening and repetition, as typical of child-directed language; Holzrichter & Meier, 2000; Pizer & Meier, 2011) when communicating about absent objects. Importantly, such differences between present and absent objects were not observed for less iconic signs, supporting the view that indeed these modifications are not just attention-grabbers. These results indicate instead that caregivers exploited the imagistic potential offered by iconicity as a resource particularly helpful in displaced contexts in helping the child map words to referents.

Using a similar design (but with the children present), Vigliocco et al. (2019) focused on child-directed spoken language, asking English speaking caregiver-child (2-3 years old) dyads to engage in conversation about toys both when they were present and absent. To better identify learning episodes, the toys were also divided into those known and unknown to the child. The authors analysed cues produced by the caregivers, coding indexical (points and object manipulations) and iconic cues (onomatopoeia and co-speech gestures) across different channels, demonstrating that these non-arbitrary multimodal behaviours are well represented in the input, accompanying almost 40% of caregiver utterances. Similarly to Perniss et al. (2017), they found that while indexical cues were overwhelmingly more common when the objects were present, onomatopoeia and iconic gestures were used more frequently in displaced contexts, and most often when the referent was unknown to the child. We illustrate examples of the different non-arbitrary cues found to constitute caregivers' input in both sign and spoken languages in *Figure 2*.
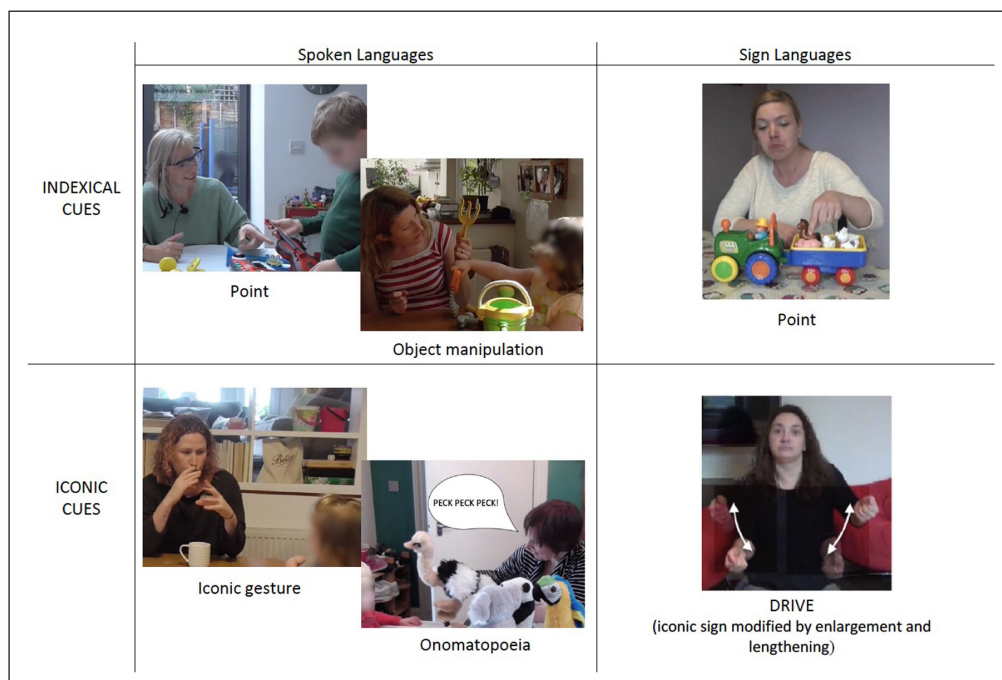

Figure 2 Examples of indexical cues (such as points and hand actions) and iconic cues (such as gestures, onomatopoeia and prosodic modulations of iconic signs) found in parental semi-naturalistic productions in spoken (Vigliocco et al., 2019) and sign languages (Perniss et al., 2017).

By giving a more comprehensive picture of how language is acquired in the whole range of possible learning contexts, Perniss et al. (2017) and Vigliocco et al. (2019) go beyond the view of non-arbitrariness as a feature present or absent in the linguistic system, asking instead what do caregivers do when they use both systemic resources and multiple channels available in face-to-face communication. In the case of Perniss et al. (2017), we propose that this approach importantly overcomes a view inherited from a first scientific view on sign languages: that few signs can be considered effectively iconic, i.e., transparent (Bellugi & Klima, 1976; Klima & Bellugi, 1979), and that iconicity necessarily declines in a language in response to structural changes (Frishberg, 1975). This view, motivated by the initial need to recognise sign languages as proper linguistic systems, equivalent to spoken languages, led researchers to emphasize their structural similarities with spoken languages, minimizing their pictorial aspects (Kendon,

2012). However, several studies have shown that the so-called *frozen* iconicity of lexical signs can be modified during language use (Brennan, 1991, 1992; Cuxac, 2003; Johnston & Schembri, 1999; Russo, 2004), showing an online process that goes in the opposite direction with respect to the system-wide diachronic tendency for forms to become less iconic. In this way, iconicity in sign languages appears as a resource that, being well-incorporated into the phonological and grammatical constraints of the system, can be also retrieved and enriched in online language use.

Vigliocco et al. (2019) show that non-arbitrariness is pervasive in spoken languages during the communication between a caregiver and a child, as they find a noticeable presence of both iconic and indexical cues in caregivers' input in both the vocal and the manual channel. Furthermore, the study shows that also in spoken languages different non-arbitrary cues are used by caregivers when they are most useful to children according to the context. Indexical cues – which signify the object by creating a visual link to it – are helpful when the referent is visually accessible: they can single it out in a cluttered visual scene. Iconic cues, on the other hand, represent the referent through a selection of its experiential properties that, even if absent from the environment, are brought into the communicative context. The study shows that iconicity is especially used when the label is unfamiliar to the child and the referred-to object is not immediately available in the environment. Here iconicity can bring to the "mind's eye" properties of the referent. These results highlight the need to take into account both different modalities and different types of learning contexts; e.g., both joint attentional and displaced, to fully understand how children (and their caregivers) exploit non-arbitrary cues as a communicative resource.

5. NON-ARBITRARINESS IN LANGUAGE PROCESSING

A role for context in language processing has long been recognized, with growing attention from researchers in the last decade or so (see reviews in Cai & Vigliocco, 2018; Meteyard & Vigliocco, 2018; Yee & Thompson-Schill, 2016). It is however the case that most current work still focuses on a single contextual factor. For example, researchers working with variants of the visual world paradigm (Tanenhaus et al., 1995) in which eye movements to depicted referents are recorded while subjects listen to speech, ignore the visual cues provided by the speaker (e.g., their gestures); those working on gestures ignore the physical context and those working on prosody often ignore all visual cues. This is done in order to secure experimental control. However, in this manner ecological validity is jeopardized (e.g., Hasson et al., 2018) and, crucially for our purposes here, visual processes based on indexical cues and imagistic processes based on iconic cues can be missed. Keeping this general issue in mind, let us briefly review some of the relevant literature.

Many studies have investigated how iconic gestures contribute to language comprehension and production. In language comprehension, iconic gestures can disambiguate lexical meaning of homonyms (Holle & Gunter, 2007), can support comprehension when the speech is degraded (Drijvers & Özyürek, 2017; Holle et al., 2010) and can provide further details useful in building situational models (e.g., if a listener hears a speaker say 'and then I paid' whilst making a writing gesture, the listener can understand that the speaker paid using a cheque; Cocks et al., 2018). Drijvers and Özyürek (2017) further showed that iconic gestures play a more important role than visible mouth patterns in supporting spoken comprehension in noise, but that both cues together can contribute to comprehension.

Gestures have also been shown to play an important role in production, for example by facilitating lexical retrieval (Hadar & Butterworth, 2009; Rauscher et al., 1996), or chunking information for verbal encoding (Alibali et al., 2000; Kita et al., 2017). Kita et al. (2007) showed that the syntactic organisation of manner and path information had an impact on the types of gestures participants produced, indicating that not only can speech and gesture affect each other, but that the semantic coordination between speech and gesture occurs online and is dependent on the linguistic context.

Iconic prosody has also been shown to contribute to meaning-making in online language use. In a series of studies, Shintel and colleagues (Shintel et al., 2006, 2014) investigated the role of prosody in both production and comprehension. Shintel, Nusbaum and Okrent (2006) showed English speaking adults animations of a dot moving in different directions (up or down) and at

different speeds (fast or slow) going left or right. When participants were asked to describe the direction of the dot, participants showed changes in F0 that matched the up/down direction of the dot's movement in the first case, and similar modulation in speech rate in the second case. Particularly in the second case, where the lexical cue does not match the prosodic modulation, the results suggest that participants convey meaning in prosody independent of the meaning conveyed lexically. Shintel, Anderson and Fenn (2014) tested whether prosodic modulations affect comprehension. They found that congruent prosody (e.g., high pitch conveying small size) allowed participants to identify referents more easily. Prosodic cues, like iconic gestures and points, contribute to meaning-making in both production and comprehension, sometimes supplying information independent of lexical content.

In the manual modality, just like in language development, the communicative context does not only provide iconic cues, but also indexical cues. A number of studies have analysed the relationship between pointing gestures and the use of demonstratives such as "this" and "that" (see Peeters & Özyürek, 2016 for a review). It has been shown that pointing is favoured over linguistic description to direct attention over small distances (Bangerter, 2004; Cooperrider, 2016), but conversely, linguistic description takes over for larger distances where pointing is less precise. Cooperrider (2016) also found that pointing could affect the type of demonstrative used, such that proximal demonstratives (e.g., "this") were preferred when points were produced, and distal demonstratives (e.g., "that") when points were absent. Taken together, these results suggest that speech and points form a combined system in face-to-face communication, used to effectively direct attention to specific locations or referents. Eye gaze can also provide a powerful visual cue to what is being talked about and facilitate processing (Holler et al., 2014).

The wide range of evidence concerning iconic and indexical communicative cues suggests that language as it is used in online, face-to-face interaction between both adults and children cannot be confined to a system of context-independent, linguistic components, nor can it be assigned only a grounding role, useful in the early stages of language learning. Rather, language being a dynamic, multimodal system situated in a given communicative and physical context, adaptively uses multiple arbitrary and non-arbitrary cues (arbitrary words and onomatopoeia available in the linguistic system, as well as iconic gesture, deixis and prosody) to contribute to context-dependent meaning making. This understanding of language as multimodal and dynamic in nature highlights the need to study different cues in combination, rather than as individual and somewhat unrelated elements of communication. A number of studies have made progress in this direction, highlighting how comprehension is modulated by multiple cues in interaction, such as gesture, eye gaze, mouth movements and prosodic modulations (Drijvers & Özyürek, 2017; Holler et al., 2014; Zhang et al., 2019). We propose that, going forward, we need language models that comprehensively examine the situated and multimodal nature of language, that account for non-arbitrary mappings from different sources, both in isolation and in combination with each other.

6. NON-ARBITRARINESS AND SITUATED LANGUAGE

We have shown in the previous sections that non-arbitrariness becomes clearly visible when we look at multimodal situated language learning and processing. In contrast, non-arbitrariness is subtle if we consider language only at the system-level. In fact, from the perspective of situated language, the extent of non-arbitrariness has not decreased to a negligible amount through evolution, as has been argued to be the case for iconicity in the linguistic system. In our corpus of semi-naturalistic language between 2-3 year old children and their caregivers, we see that approximately 39% of the utterances produced by caregivers contain indexical cues, and 17%[1] iconic cues (Vigliocco et al., 2019). Thus, far from being negligible, non-arbitrary multimodal cues are pervasive in communication to young children. Why would this be the case? For iconicity, as already mentioned above, we argue that it can have an important role in providing imagistic information about the objects or events being discussed, helping in bringing back to mind the memory traces of an already experienced object or providing cues to build the conceptual knowledge relative to a completely new object (Vigliocco et al., 2019). A second important observation from our corpus is that while iconicity in speech (in the form of

---

1   This estimate does not include iconic modulations of prosody, hence it is bound to be under-estimating the actual incidence.

onomatopoeia) *decreases* significantly between 2 to 4 years of age, it does *increase* significantly in iconic gestures used by caregivers (Motamedi et al., 2020). This fact suggests that rather than disappearing, iconicity moves from the linguistic form, to leave room for a larger linguistic repertoire, to context-dependent recruitment of other modalities. Note here that in sign languages, the possibility to offload iconicity to the auditory channel is unavailable, which may explain why sign languages in general tend to have far more context-independent iconicity in the primary channel than spoken languages; i.e., meaningful form-meaning association in the linguistic system, for example at the phonological or lexical level. With respect to indexicality, we show that this strategy, already known to be used very early on by caregivers and children, is pervasive in joint-attentional communication through the use of different cues like pointing gestures as well as object manipulations which are able to anchor the message to the physical context increasing its referential transparency (Vigliocco et al., 2019).

For children and adults, non-arbitrariness in the communicative context is present beyond the negligeable remnants we see in linguistic forms. Although we do not have at the moment any estimate of how often multimodal cues would be used in language processing by adults, in the previous sections we presented abundant evidence for their presence and for their facilitatory function in processing. We argue that the multimodal communicative cues present in situated language processing serve the key function of supporting the development of an aligned situational model shared by the speaker/signer and the addressee. For the addressee, non-arbitrary cues produced by the speaker/signer support this function in at least two ways. First, by reducing potential ambiguity in the developing situational model. They provide direct visual links (with indexicality) and imagistic links (with iconicity) that can clarify and complement the primary channel; e.g., points can direct the attention to the referred-to object or a specific part of it, disambiguating the message; iconic gestures can provide information about the direction of motion in a 3D model that can be absent in speech; prosody and sound effects can replace words in providing information about speed of events, acoustic features of events, etc. Second and more broadly, they can reduce the cognitive effort required in generating situational models when the physical context does not provide any cue (when the language is displaced) and therefore the model needs to be generated internally. As displacement is a key feature of language (Hockett, 1960; Perniss & Vigliocco, 2014), it is not surprising that our communicative system would maintain features such as iconicity that can make displaced language easier to process.

For speakers and signers, multimodal non-arbitrary cues may also serve multiple functions. For example, for iconic gestures Kita et al. (2017) proposed the gesture-for-conceptualization hypothesis, according to which iconic gestures affect speakers' cognitive processes in four different manners: they activate, manipulate, package, and explore spatio-motoric information for thinking and speaking. These four functions are shaped by gesture's ability to schematize information. Thus, iconic gestures would serve self-oriented functions in addition to communicative functions. Onomatopoeia have been shown to be more resistant to brain damage than arbitrary words (Meteyard et al., 2015) possibly because retrieval of their phonological form can benefit from the support of non-linguistic information (see Vigliocco & Kita, 2006). Thus, multimodal iconicity in situated language could bring benefits not just to addresses but also speakers.

7. CONCLUSIONS

In this paper we provide a novel theoretical view on non-arbitrariness and its role in language which incorporates the language as a system perspective — according to which language is conceived as a system of symbols and rules that evolved at a population level — into the wider view of language as situated. Under this view, language — in its prime form i.e., face-to-face communication — is always embedded in a physical and communicative context in which speakers can and do use non-arbitrary multimodal cues, plausibly at least in part to support fluent and effortful production, and in which addressees can use these cues to learn language and to develop a situational model aligned with the speaker/signer's. Crucially, the body of evidence we have reviewed suggests that cues beyond just speech and signs which are present in the face-to-face interactive context contribute to language acquisition and processing. Thus, language should be investigated as a multiplex and context dependent phenomenon if we seek a full account of how language has evolved, is learnt and is used in the real-world.

ETHICS AND CONSENT

Informed consent for children participating in studies reported in Vigliocco et al. (2019) and Motamedi et al. (2020) was given by caregivers on their behalf. These studies were approved by the UCL ethics committees (Ecological Language (ECOLANG): 0143/003, The role of iconicity in language development: 0143/002). We reproduce images from three studies cited in the manuscript (Motamedi et al., 2020; Perniss et al., 2017; Vigliocco et al., 2019). Permission to reprint images from Perniss et al. (2017) was obtained from John Wiley & Sons publications.

ACKNOWLEDGEMENTS

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHOR CONTRIBUTIONS

All authors contributed to conception of ideas, writing and editing of the manuscript.

AUTHOR AFFILIATIONS

Margherita Murgiano
Experimental Psychology, University College London, GB

Yasamin Motamedi
Experimental Psychology, University College London, GB

Gabriella Vigliocco (iD) *orcid.org/0000-0002-7190-3659*
Experimental Psychology, University College London, GB

REFERENCES

Abner, N., Cooperrider, K., & **Goldin-Meadow, S.** (2015). Gesture for Linguists: A Handy Primer. *Language and Linguistics Compass, 9*(11), 437–449. DOI: *https://doi.org/10.1111/lnc3.12168*

Alibali, M. W., Kita, S., & **Young, A. J.** (2000). Gesture and the process of speech production: We think, therefore we gesture. *Language and Cognitive Processes, 15*(6), 593–613. DOI: *https://doi.org/10.1080/016909600750040571*

Alibali, M. W., Mitchell, N., & **Fujimori, Y.** (2011). Gesture in the mathematics classroom: What's the point? In N. L. Stein & S. W. Raudenbush (Eds.), *Developmental Cognitive Science Goes to School* (pp. 219–234). Routledge.

Austin, J. L. (1962). *How to do things with words*. Oxford University Press.

Bangerter, A. (2004). Using Pointing and Describing to Achieve Joint Focus of Attention in Dialogue. *Psychological Science, 15*(6), 415–419. DOI: *https://doi.org/10.1111/j.0956-7976.2004.00694.x*

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology, 59,* 617–645. DOI: *https://doi.org/10.1146/annurev.psych.59.103006.093639*

Bellugi, U., & **Klima, E. S.** (1976). Two Faces of Sign: Iconic and Abstract*. *Annals of the New York Academy of Sciences, 280*(1), 514–538. DOI: *https://doi.org/10.1111/j.1749-6632.1976.tb25514.x*

Bohn, M., & **Frank, M. C.** (2019). The Pervasive Role of Pragmatics in Early Language. *Annual Review of Developmental Psychology, 1*(1), 223–249. DOI: *https://doi.org/10.1146/annurev-devpsych-121318-085037*

Bohn, M., & **Köymen, B.** (2018). Common Ground and Development. *Child Development Perspectives, 12*(2), 104–108. DOI: *https://doi.org/10.1111/cdep.12269*

Boyes-Braem, P. (1981). *Features of the handshape in American Sign Language* [Ph.D.]. Berkeley: University of California.

Boyes-Braem, P. (1986). Two aspects of pyscholinguistic research: Iconicity and temporal structure. In B. T. Tervoort (Ed.), *Signs of life: Proceedings of the second european congress on sign language research, 50,* 65–74. Publication of the Institute for General Linguistics, University of Amsterdam.

Brennan, M. (1991). Making borrowings work in British Sign Language. In D. Brentari (Ed.), *Foreign vocabulary in sign languages: Cross-linguistic investigation of word formation, Sign* (pp. 49–86). Lawrence Erlbaum.

Brennan, M. (1992). The visual world of British Sign Language: An introduction. In D. Brien (Ed.), *Dictionary of British Sign Language/English*. Faber and Faber.

Brentari, D. (2012). Phonology. In R. Pfau, M. Steinbach, & B. Woll (Eds.), *Sign Language: An International Handbook*. De Gruyter Mouton. DOI: *https://doi.org/10.1515/9783110261325.21*

Cai, Z., & **Vigliocco, G.** (2018). Word Processing. In J. T. Wixted & S. L. Thompson-Schill (Eds.), *Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience: Vols 3. Language and Thought* (4th ed.). DOI: *https://doi.org/10.1002/9781119170174.epcn303*

Cartmill, E. A., Armstrong, B. F., Gleitman, L. R., Goldin-Meadow, S., Medina, T. N., & **Trueswell, J. C.** (2013). Quality of early parent input predicts child vocabulary 3 years later. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(28), 11278–83. DOI: *https://doi.org/10.1073/pnas.1309518110*

Caselli, N. K., & **Pyers, J. E.** (2019). Degree and not type of iconicity affects sign language vocabulary acquisition. *Journal of Experimental Psychology. Learning, Memory, and Cognition*. DOI: *https://doi.org/10.1037/xlm0000713*

Chandrasekaran, C., Trubanova, A., Stillittano, S., Caplier, A., & **Ghazanfar, A. A.** (2009). The Natural Statistics of Audiovisual Speech. *PLOS Computational Biology*, *5*(7), e1000436. DOI: *https://doi.org/10.1371/journal.pcbi.1000436*

Christensen, P., Fusaroli, R., & **Tylén, K.** (2016). Environmental constraints shaping constituent order in emerging communication systems: Structural iconicity, interactive alignment and conventionalization. *Cognition*, *146*, 67–80. DOI: *https://doi.org/10.1016/j.cognition.2015.09.004*

Clark, H. H. (1996). *Using language*. Cambridge University Press. DOI: *https://doi.org/10.1017/CBO9780511620539*

Clark, H. H. (2016). Depicting as a Method of Communication. *Psychological Review*, *123*(3), 324–347. DOI: *https://doi.org/10.1037/rev0000026*

Cocks, N., Byrne, S., Pritchard, M., Morgan, G., & **Dipper, L.** (2018). Integration of speech and gesture in aphasia. *International Journal of Language & Communication Disorders*, *53*(3), 584–591. DOI: *https://doi.org/10.1111/1460-6984.12372*

Cooperrider, K. (2016). The Co-Organization of Demonstratives and Pointing Gestures. *Discourse Processes*, *53*(8), 632–656. DOI: *https://doi.org/10.1080/0163853X.2015.1094280*

Cormier, K., Fenlon, J., & **Schembri, A.** (2015). Indicating verbs in British Sign Language favour motivated use of space. *Open Linguistics*, *1*, 684–707. DOI: *https://doi.org/10.1515/opli-2015-0025*

Cutler, A., Dahan, D., & **van Donselaar, W.** (1997). Prosody in the Comprehension of Spoken Language: A Literature Review: *Language and Speech*, *40*(2). DOI: *https://doi.org/10.1177/002383099704000203*

Cutler, A., Mehler, J., Norris, D., & **Segui, J.** (1986). The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language*, *25*(4), 385–400. DOI: *https://doi.org/10.1016/0749-596X(86)90033-1*

Cuxac, C. (2003). Langue et langage: Un apport critique de la langue des signes française. *Langue française*, *137*(1), 12–31. DOI: *https://doi.org/10.3406/lfr.2003.1054*

De Ruiter, L. E., Theakston, A. L., Brandt, S., & **Lieven, E. V. M.** (2018). Iconicity affects children's comprehension of complex sentences: The role of semantics, clause order, input and individual differences. *Cognition*, *171*(March 2017), 202–224. DOI: *https://doi.org/10.1016/j.cognition.2017.10.015*

de Saussure, F. (1916). *Cours de linguistique générale*. Payot.

D'Entremont, B. (2000). A perceptual–attentional explanation of gaze following in 3- and 6-month-olds. *Developmental Science*, *3*(3), 302–311. DOI: *https://doi.org/10.1111/1467-7687.00124*

Diessel, H. (2008). Iconicity of sequence: A corpus-based analysis of the positioning of temporal adverbial clauses in English. *Cognitive Linguistics*, *19*(3), 465–490. DOI: *https://doi.org/10.1515/COGL.2008.018*

Dingemanse, M. (2018). Redrawing the margins of language: Lessons from research on ideophones. *Glossa: A Journal of General Linguistics*, *3*(1), 4. DOI: *https://doi.org/10.5334/gjgl.444*

Dingemanse, M., Blasi, D. E., Lupyan, G., Christiansen, M. H., & **Monaghan, P.** (2015). Arbitrariness, iconicity and systematicity in language. *Trends in Cognitive Sciences*, *19*(10), 603–615. DOI: *https://doi.org/10.1016/j.tics.2015.07.013*

Dingemanse, M., Schuerman, W., Reinisch, E., Tufvesson, S., & **Mitterer, H.** (2016). What sound symbolism can and cannot do: Testing the iconicity of ideophones from five languages. *Language*, *92*(2), e117–e133. DOI: *https://doi.org/10.1353/lan.2016.0034*

Drijvers, L., & **Özyürek, A.** (2017). Visual Context Enhanced: The Joint Contribution of Iconic Gestures and Visible Speech to Degraded Speech Comprehension. *Journal of Speech, Language, and Hearing Research*, *60*(1), 212–222. DOI: *https://doi.org/10.1044/2016_JSLHR-H-16-0101*

Emmorey, K. (2014). Iconicity as structure mapping. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *369*(1651), 20130301. DOI: *https://doi.org/10.1098/rstb.2013.0301*

Emmorey, K., & **Herzig, M.** (2003). Categorical Versus Gradient Properties of Classifier Constructions in ASL. In K. Emmorey (Ed.), *Perspectives on Classifier Constructions in Sign Languages* (pp. 221–246). Lawrence Erlbaum. DOI: *https://doi.org/10.4324/9781410607447*

Fay, N., Arbib, M., & **Garrod, S.** (2013). How to bootstrap a human communication system. *Cognitive Science*, *37*(7), 1356–67. DOI: *https://doi.org/10.1111/cogs.12048*

Fontana, S. (2008). *Mouth actions as gesture in sign language* [Text]. *https://doi.org/info:doi/10.1075/gest.8.1.08fon*

Frishberg, N. (1975). Arbitrariness and Iconicity: Historical Change in American Sign Language. *Language*, *51*(3), 696–719. DOI: *https://doi.org/10.2307/412894*

Garrod, S., Fay, N., Lee, J., Oberlander, J., & **MacLeod, T.** (2007). Foundations of representation: Where might graphical symbol systems come from? *Cognitive Science*, *31*(6), 961–987. DOI: *https://doi.org/10.1080/03640210701703659*

Gasser, M. (2004). The origins of arbitrariness in language. *Proceedings of the 26th Annual Conference of the Cognitive Science Society*, 4–7.

Goodrich, W., & **Hudson Kam, C. L.** (2009). Co-speech gesture as input in verb learning. *Developmental Science*, *12*(1), 81–87. DOI: *https://doi.org/10.1111/j.1467-7687.2008.00735.x*

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Speech Acts* (pp. 41–58). Academic Press. DOI: *https://doi.org/10.1163/9789004368811_003*

Hadar, U., & **Butterworth, B.** (2009). Iconic gestures, imagery, and word retrieval in speech. *Semiotica*, *115*(1–2), 147–172. DOI: *https://doi.org/10.1515/semi.1997.115.1-2.147*

Hasson, U., Egidi, G., Marelli, M., & **Willems, R. M.** (2018). Grounding the neurobiology of language in first principles: The necessity of non-language-centric explanations for language comprehension. *Cognition*, *180*, 135–157. DOI: *https://doi.org/10.1016/j.cognition.2018.06.018*

Herold, D. S., Nygaard, L. C., Chicos, K. A., & **Namy, L. L.** (2011). The developing role of prosody in novel word interpretation. *Journal of Experimental Child Psychology*, *108*(2), 229–241. DOI: *https://doi.org/10.1016/j.jecp.2010.09.005*

Herold, D. S., Nygaard, L. C., & **Namy, L. L.** (2011). Say It Like You Mean It: Mothers' Use of Prosody to Convey Word Meaning: *Language and Speech*. DOI: *https://doi.org/10.1177/0023830911422212*

Hockett, C. F. (1960). The origin of speech. *Scientific American*, *203*, 88–96. DOI: *https://doi.org/10.1038/scientificamerican0960-88*

Hofsten, C., Dahlström, E., & **Fredriksson, Y.** (2005). 12-Month-Old Infants' Perception of Attention Direction in Static Video Images. *Infancy*, *8*(3), 217–231. DOI: *https://doi.org/10.1207/s15327078in0803_2*

Holle, H., & **Gunter, T. C.** (2007). The Role of Iconic Gestures in Speech Disambiguation: ERP Evidence. *Journal of Cognitive Neuroscience*, *19*(7), 1175–1192. DOI: *https://doi.org/10.1162/jocn.2007.19.7.1175*

Holle, H., Obleser, J., Rueschemeyer, S.-A., & **Gunter, T. C.** (2010). Integration of iconic gestures and speech in left superior temporal areas boosts speech comprehension under adverse listening conditions. *NeuroImage*, *49*(1), 875–884. DOI: *https://doi.org/10.1016/j.neuroimage.2009.08.058*

Holler, J., & **Levinson, S. C.** (2019). Multimodal Language Processing in Human Communication. *Trends in Cognitive Sciences*, *23*(8), 639–652. DOI: *https://doi.org/10.1016/j.tics.2019.05.006*

Holler, J., Schubotz, L., Kelly, S., Hagoort, P., Schuetze, M., & **Özyürek, A.** (2014). Social eye gaze modulates processing of speech and co-speech gesture. *Cognition*, *133*(3), 692–697. DOI: *https://doi.org/10.1016/j.cognition.2014.08.008*

Holzrichter, A. S., & **Meier, R. P.** (2000). Child-directed signing in American Sign Language. In *Language acquisition by eye* (pp. 25–40). Lawrence Erlbaum Associates Publishers.

Imai, M., & **Kita, S.** (2014). The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *369*(1651), 20130298–. DOI: *https://doi.org/10.1098/rstb.2013.0298*

Imai, M., Kita, S., Nagumo, M., & **Okada, H.** (2008). Sound symbolism facilitates early verb learning. *Cognition*, *109*(1), 54–65. DOI: *https://doi.org/10.1016/j.cognition.2008.07.015*

Iverson, J. M., Capirci, O., Longobardi, E., & **Caselli, M. C.** (1999). Gesture in mother-child interactions. *Cognitive Development*, *14*, 57–75. DOI: *https://doi.org/10.1016/S0885-2014(99)80018-5*

Iwasaki, N., Vinson, D. P., & **Vigliocco, G.** (2007). What do English Speakers Know about gera-gera and yota-yota?: A Cross-linguistic Investigation of Mimetic Words for Laughing and Walking. *Japanese-Language Education around the Globe*, *17*(6), 53–78.

Johnston, T., & **Schembri, A. C.** (1999). On Defining Lexeme in a Signed Language. *Sign Language & Linguistics*, *2*(2), 115–185. DOI: *https://doi.org/10.1075/sll.2.2.03joh*

Kelly, S., Healey, M., Özyürek, A., & **Holler, J.** (2015). The processing of speech, gesture, and action during language comprehension. *Psychonomic Bulletin & Review*, *22*(2), 517–523. DOI: *https://doi.org/10.3758/s13423-014-0681-7*

Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge University Press. DOI: *https://doi.org/10.1017/CBO9780511807572*

Kendon, A. (2008). Some reflections on the relationship between 'gesture' and 'sign'. *Gesture*, *8(3)*, 348–366. DOI: *https://doi.org/10.1075/gest.8.3.05ken*

Kendon, A. (2012). Language and kinesic complexity: Reflections on "Dedicated gestures and the emergence of sign language" by Wendy Sandler. *Gesture, 12*(3), 308–326. DOI: *https://doi.org/10.1075/gest.12.3.02ken*

Kendon, A. (2014). Semiotic diversity in utterance production and the concept of 'language'. *Philosophical Transactions of the Royal Society B: Biological Sciences, 369*(1651), 20130293. DOI: *https://doi.org/10.1098/rstb.2013.0293*

Kita, S., Alibali, M. W., & **Chu, M.** (2017). How do gestures influence thinking and speaking? The gesture-for-conceptualization hypothesis. *Psychological Review, 124*(3), 245–266. DOI: *https://doi.org/10.1037/rev0000059*

Kita, S., Özyürek, A., Allen, S., Brown, A., Furman, R., & **Ishizuka, T.** (2007). Relations between syntactic encoding and co-speech gestures: Implications for a model of speech and gesture production. *Language and Cognitive Processes, 22*(8), 1212–1236. DOI: *https://doi.org/10.1080/01690960701461426*

Klima, E. S., & **Bellugi, U.** (1979). *The Signs of Language*. Harvard University Press.

Knoeferle, K., Li, J., Maggioni, E., & **Spence, C.** (2017). What drives sound symbolism? Different acoustic cues underlie sound-size and sound-shape mappings. *Nature, 7*(1), 5562. DOI: *https://doi.org/10.1038/s41598-017-05965-y*

Köhler, W. (1929). *Gestalt psychology*. Liveright.

Laing, C. E. (2014). A phonological analysis of onomatopoeia in early word production. *First Language, 34*(5), 387–405. DOI: *https://doi.org/10.1177/0142723714550110*

Levinson, S. C., & **Holler, J.** (2014). The origin of human multi-modal communication. *Philosophical Transactions of the Royal Society B: Biological Sciences, 369*(1651), 20130302. DOI: *https://doi.org/10.1098/rstb.2013.0302*

Liddell, S. K. (2003). *Grammar, gesture and meaning in American Sign Language*. Cambridge University Press. DOI: *https://doi.org/10.1017/CBO9780511615054*

Little, H., Eryılmaz, K., & **de Boer, B.** (2017). Signal dimensionality and the emergence of combinatorial structure. *Cognition, 168*, 1–15. DOI: *https://doi.org/10.1016/j.cognition.2017.06.011*

Lockwood, G., Dingemanse, M., & **Hagoort, P.** (2016). Sound-symbolism boosts novel word learning. *Journal of Experimental Psychology: Learning Memory and Cognition, 42*(8), 1274–1281. DOI: *https://doi.org/10.1037/xlm0000235*

Lupyan, G., & **Winter, B.** (2018). Language is more abstract than you think, or, why aren't languages more iconic? *Philosophical Transactions of the Royal Society B: Biological Sciences, 373*(1752), 20170137. DOI: *https://doi.org/10.1098/rstb.2017.0137*

Marchand, H. (1959). Phonetic symbolism in English word-formation. *Indogermanische Forschungen, 64*, 256–277. DOI: *https://doi.org/10.1515/9783110243062.256*

Mcgurk, H., & **Macdonald, J.** (1976). Hearing lips and seeing voices. *Nature, 264*(5588), 746–748. DOI: *https://doi.org/10.1038/264746a0*

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

Meteyard, L., Stoppard, E., Snudden, D., Cappa, S. F., & **Vigliocco, G.** (2015). When semantics aids phonology: A processing advantage for iconic word forms in aphasia—ScienceDirect. *Neuropsychologia, 76*, 264–275. DOI: *https://doi.org/10.1016/j.neuropsychologia.2015.01.042*

Meteyard, L., & **Vigliocco, G.** (2018). Lexico-Semantics. *The Oxford Handbook of Psycholinguistics*. DOI: *https://doi.org/10.1093/oxfordhb/9780198786825.013.4*

Mohr, S. (2014). *Mouth Actions in Sign Languages: An empirical study of Irish Sign Language*. De Gruyter. DOI: *https://doi.org/10.1515/9781614514978*

Mondada, L. (2019). Contemporary issues in conversation analysis: Embodiment and materiality, multimodality and multisensoriality in social interaction. *Journal of Pragmatics, 145*, 47–62. DOI: *https://doi.org/10.1016/j.pragma.2019.01.016*

Motamedi, Y., Murgiano, M., Perniss, P., Grzyb, B., Brieke, R., Gu, Y., Marshall, C. R., Wonnacott, E., & **Vigliocco, G.** (2020). *There is much more than just words in caregivers' communication to children*. Manuscript in preparation.

Mumford, K. H., & **Kita, S.** (2014). Children Use Gesture to Interpret Novel Verb Meanings. *Child Development, 85*(3), 1181–1189. DOI: *https://doi.org/10.1111/cdev.12188*

Namy, L. L., Campbell, A. L., & **Tomasello, M.** (2004). The changing role of iconicity in non-verbal symbol learning: A U-shaped trajectory in the acquisition of arbitrary gestures. *Journal of Cognition and Development, 5*(1), 37–57. DOI: *https://doi.org/10.1207/s15327647jcd0501_3*

Nygaard, L. C., Herold, D. S., & **Namy, L. L.** (2009). The semantics of prosody: Acoustic and perceptual evidence of prosodic correlates to word meaning. *Cognitive Science, 33*(1), 127–146. DOI: *https://doi.org/10.1111/j.1551-6709.2008.01007.x*

Oda, H. (2000). *An embodied semantic mechanism for mimetic words in Japanese* [Ph.D., Indiana University]. *https://search.proquest.com/docview/304600174/abstract/65B694AB3D32419BPQ/1*

O'Neill, M., Bard, K. A., Linnell, M., & **Fluck, M.** (2005). Maternal gestures with 20-month-old infants in two contexts. *Developmental Science, 8*(4), 352–359. DOI: *https://doi.org/10.1111/j.1467-7687.2005.00423.x*

Ortega, G. (2017). Iconicity and Sign Lexical Acquisition: A Review. *Frontiers in Psychology, 8*. DOI: *https://doi.org/10.3389/fpsyg.2017.01280*

Peeters, D., & **Özyürek, A.** (2016). This and That Revisited: A Social and Multimodal Approach to Spatial Demonstratives. *Frontiers in Psychology, 7*. DOI: *https://doi.org/10.3389/fpsyg.2016.00222*

Peirce, C. S. (1974). *Collected papers of Charles Sanders Pierce*. Harvard University Press.

Perlman, M., Dale, R., & **Lupyan, G.** (2015). Iconicity can ground the creation of vocal symbols. *Royal Society Open Science, 2*(8), 150152. DOI: *https://doi.org/10.1098/rsos.150152*

Perniss, P. (2018). Why We Should Study Multimodal Language. *Frontiers in Psychology, 9*(June), 1–5. DOI: *https://doi.org/10.3389/fpsyg.2018.01109*

Perniss, P., Lu, J. C., Morgan, G., & **Vigliocco, G.** (2017). Mapping language to the world: The role of iconicity in the sign language input. *Developmental Science, July 2015*, 1–23. DOI: *https://doi.org/10.1111/desc.12551*

Perniss, P., Thompson, R. L., & **Vigliocco, G.** (2010). Iconicity as a general property of language: Evidence from spoken and signed languages. *Frontiers in Psychology, 1*(DEC), 1–15. DOI: *https://doi.org/10.3389/fpsyg.2010.00227*

Perniss, P., & **Vigliocco, G.** (2014). The bridge of iconicity: From a world of experience to the experience of language. *Philosophical Transactions of the Royal Society B: Biological Sciences, 369*. DOI: *https://doi.org/10.1098/rstb.2013.0300*

Perry, L. K., Perlman, M., & **Lupyan, G.** (2015). Iconicity in english and Spanish and its relation to lexical category and age of acquisition. *PLoS ONE, 10*(9), 1–17. DOI: *https://doi.org/10.1371/journal.pone.0137147*

Perry, L. K., Perlman, M., Winter, B., Massaro, D. W., & **Lupyan, G.** (2017). Iconicity in the speech of children and adults. *Developmental Science, 21*(3), 1–8. DOI: *https://doi.org/10.1111/desc.12572*

Pfau, R., & **Quer, J.** (2010). Nonmanuals: Their grammatical and prosodic roles. In D. Brentari (Ed.), *Sign Languages* (pp. 381–402). Cambridge University Press. DOI: *https://doi.org/10.1017/CBO9780511712203.018*

Pietrandrea, P. (2002). Iconicity and Arbitrariness in Italian Sign Language. *Sign Language Studies, 2*(3), 296–321. DOI: *https://doi.org/10.1353/sls.2002.0012*

Pizer, G., & **Meier, R. P.** (2011). Child-directed signing as a linguistic register. In R. Channon & H. van der Hulst (Eds.), *Formational units in signed languages* (pp. 65–83). De Gruyter Mouton. DOI: *https://doi.org/10.1515/9781614510680.65*

Quine, W. V. O. (1960). *Word and Object*. MIT Press.

Rączaszek-Leonardi, J., Nomikou, I., Rohlfing, K. J., & **Deacon, T. W.** (2018). Language Development From an Ecological Perspective: Ecologically Valid Ways to Abstract Symbols. *Ecological Psychology, 30*(1), 39–73. DOI: *https://doi.org/10.1080/10407413.2017.1410387*

Ramachandran, V. S., & **Hubbard, E. M.** (2001). Synaesthesia—A window into perception, thought and language. *Journal of Consciousness Studies, 8*(12), 3–34.

Rauscher, F. H., Krauss, R. M., & **Chen, Y.** (1996). Gesture, Speech, and Lexical Access: The Role of Lexical Movements in Speech Production. *Psychological Science, 7*(4), 226–231. DOI: *https://doi.org/10.1111/j.1467-9280.1996.tb00364.x*

Roberts, G., Lewandowski, J., & **Galantucci, B.** (2015). How communication changes when we cannot mime the world: Experimental evidence for the effect of iconicity on combinatoriality. *Cognition, 141*, 52–66. DOI: *https://doi.org/10.1016/j.cognition.2015.04.001*

Rowe, M. L., & **Goldin-Meadow, S.** (2009). Differences in Early Gesture Explain SES Disparities in Child Vocabulary Size at School Entry. *Science, 323*(5916), 951–953. DOI: *https://doi.org/10.1126/science.1167025*

Russo, T. (2004). *La mappa poggiata sull'isola. Iconicità e metafora nelle lingue dei segni e nelle lingue vocali*. Centro Editoriale e Librario Università degli Studi della Calabria.

Sallandre, M.-A., & **Cuxac, C.** (2001). Iconicity in sign language: A theoretical and methodological point of view. *Proceedings of the International Gesture Workshop*, 171–180.

Sandler, W. (1999). The Medium and the Message: Prosodic Interpretation of Linguistic Content in Israeli Sign Language. *Sign Language & Linguistics, 2*(2), 187–215. DOI: *https://doi.org/10.1075/sll.2.2.04san*

Searle, J. (1969). *Speech Acts: An essay in the philosophy of language*. Cambridge University Press. DOI: *https://doi.org/10.1017/CBO9781139173438*

Senju, A., & **Csibra, G.** (2008). Gaze Following in Human Infants Depends on Communicative Signals. *Current Biology, 18*(9), 668–671. DOI: *https://doi.org/10.1016/j.cub.2008.03.059*

Shintel, H., Anderson, N. L., & **Fenn, K. M.** (2014). Talk This Way: The Effect of Prosodically Conveyed Semantic Information on Memory for Novel Words. *Journal of Experimental Psychology General, 143*(4), 1437–1442. DOI: *https://doi.org/10.1037/a0036605*

Shintel, H., Nusbaum, H. C., & **Okrent, A.** (2006). Analog acoustic expression in speech communication. *Journal of Memory and Language, 55*(2), 167–177. DOI: *https://doi.org/10.1016/j.jml.2006.03.002*

Sidhu, D. M., & **Pexman, P. M.** (2018). Lonely sensational icons: Semantic neighbourhood density, sensory experience and iconicity. *Language, Cognition and Neuroscience, 33*(1), 25–31. DOI: *https://doi.org/10.1080/23273798.2017.1358379*

Sidhu, D. M., Vigliocco, G., & **Pexman, P. M.** (2020). Effects of iconicity in lexical decision. *Language and Cognition*, *12*, 1–18. DOI: *https://doi.org/10.1017/langcog.2019.36*

Slobin, D. I. (2008). Breaking the Molds: Signed Languages and the Nature of Human Language. *Sign Language Studies*, *8*(2), 114–130. DOI: *https://doi.org/10.1353/sls.2008.0004*

Snedeker, J., & **Gleitman, L. R.** (2004). Why It Is Hard to Label Our Concepts. In *Weaving a lexicon* (pp. 257–293). MIT Press.

Staudte, M., & **Crocker, M. W.** (2011). Investigating joint attention mechanisms through spoken human–robot interaction. *Cognition*, *120*(2), 268–291. DOI: *https://doi.org/10.1016/j.cognition.2011.05.005*

Stec, K., Huiskes, M., & **Redeker, G.** (2016). Multimodal quotation: Role shift practices in spoken narratives. *Journal of Pragmatics*, *104*, 1–17. DOI: *https://doi.org/10.1016/j.pragma.2016.07.008*

Strickland, B., Geraci, C., Chemla, E., Schlenker, P., Kelepir, M., & **Pfau, R.** (2015). Event representations constrain the structure of language: Sign language as a window into universally accessible linguistic biases. *Proceedings of the National Academy of Sciences*, *21*, 1–6. DOI: *https://doi.org/10.1073/pnas.1423080112*

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & **Sedivy, J. C.** (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*(5217), 1632–1634. DOI: *https://doi.org/10.1126/science.7777863*

Tardif, T., Fletcher, P., Liang, W., Zhang, Z., Kaciroti, N., & **Marchman, V. A.** (2008). Baby's first 10 words. *Developmental Psychology*, *44*(4), 929–938. DOI: *https://doi.org/10.1037/0012-1649.44.4.929*

Theisen, C. A., Oberlander, J., & **Kirby, S.** (2010). Systematicity and arbitrariness in novel communication systems. *Interaction Studies*, *11*(1), 14–32. DOI: *https://doi.org/10.1075/is.11.1.08the*

Thompson, R. L., Vinson, D. P., & **Vigliocco, G.** (2010). The link between form and meaning in British sign language: Effects of iconicity for phonological decisions. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *36*(4), 1017–1027. DOI: *https://doi.org/10.1037/a0019339*

Tolar, T. D., Lederberg, A. R., Gokhale, S., & **Tomasello, M.** (2008). The development of the ability to recognize the meaning of iconic signs. *Journal of Deaf Studies and Deaf Education*, *13*(2), 225–240. DOI: *https://doi.org/10.1093/deafed/enm045*

Tomasello, M. (2008). *The Origins of Human Communication*. MIT Press. DOI: *https://doi.org/10.7551/mitpress/7551.001.0001*

Tomasello, M., & **Barton, M.** (1994). Learning Words in Nonostensive Contexts. *Developmental Psychology*, *30*(5), 639–650. DOI: *https://doi.org/10.1037/0012-1649.30.5.639*

Tomasello, M., & **Kruger, A. C.** (1992). Joint attention on actions: Acquiring verbs in ostensive and non-ostensive contexts*. *Journal of Child Language*, *19*(2), 311–333. DOI: *https://doi.org/10.1017/S0305000900011430*

Tomasello, M., Strosberg, R., & **Akhtar, N.** (1996). Eighteen-month-old children learn words in non-ostensive contexts. *Journal of Child Language*, *23*(1), 157–176. DOI: *https://doi.org/10.1017/S0305000900010138*

Vigliocco, G., & **Kita, S.** (2006). Language-specific properties of the lexicon: Implications for learning and processing. *Language and Cognitive Processes*, *21*(7–8), 790–816. DOI: *https://doi.org/10.1080/016909600824070*

Vigliocco, G., Motamedi, Y., Murgiano, M., Wonnacott, E., Marshall, C. R., Milan Maillo, I., & **Perniss, P.** (2019). Onomatopoeias, gestures, actions and words in the input to children: How do caregivers use multimodal cues in their communication to children? *Proceedings of the 41st Annual Conference of the Cognitive Science Society*. DOI: *https://doi.org/10.31234/osf.io/v263k*

Vigliocco, G., Perniss, P., & **Vinson, D.** (2014). Language as a multimodal phenomenon: Implications for language learning, processing and evolution. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *369*(1651), 20130292. DOI: *https://doi.org/10.1098/rstb.2013.0292*

Vinson, D. P., Cormier, K., Denmark, T., Schembri, A., & **Vigliocco, G.** (2008). The British Sign Language (BSL) norms for age of acquisition, familiarity, and iconicity. *Behavior Research Methods*, *40*(4), 1079–1087. DOI: *https://doi.org/10.3758/BRM.40.4.1079*

Vinson, D., Thompson, R. L., Skinner, R., & **Vigliocco, G.** (2015). A faster path between meaning and form? Iconicity facilitates sign recognition and production in British Sign Language. *Journal of Memory and Language*, *82*, 56–85. DOI: *https://doi.org/10.1016/j.jml.2015.03.002*

Wilbur, R. B. (2004). Complex predicates involving events, time and aspect: Is this why sign languages look so similar? In J. Quer (Ed.), *Signs of the time: Selected papers from TISLR 2004*. Signum Press.

Winter, B., Perlman, M., Perry, L. K., & **Lupyan, G.** (2017). Which words are most iconic? *Interaction Studies*, *18*(3), 430–451. DOI: *https://doi.org/10.1075/is.18.3.07win*

Woll, B. (2014). Moving from hand to mouth: Echo phonology and the origins of language. *Name: Frontiers in Psychology*. DOI: *https://doi.org/10.3389/fpsyg.2014.00662*

Yee, E., & **Thompson-Schill, S. L.** (2016). Putting concepts into context. *Psychonomic Bulletin & Review*, *23*(4), 1015–1027. DOI: *https://doi.org/10.3758/s13423-015-0948-7*

Zhang, Y., Frassinelli, D., Tuomainen, J., Skipper, J. I., & **Vigliocco, G.** (2019). More than words: The online orchestration of word predictability, prosody, gesture, and mouth movements during natural language comprehension. *Manuscript Submitted for Publication*. DOI: *https://doi. org/10.1101/2020.01.08.896712*