The role of iconicity and simultaneity in efficient communication in the visual modality: evidence from LIS (Italian Sign Language)

Anita Slonimska

Published by
LOT
Binnengasthuisstraat 9
1012 ZA Amsterdam
The Netherlands

phone: +31 20 525 2461

e-mail: lot@uva.nl
http://www.lotschool.nl

ISBN: 978-94-6093-413-1
DOI: https://dx.medra.org/10.48273/LOT0630
NUR: 616

The role of iconicity and simultaneity in efficient communication in the visual modality: evidence from LIS (Italian Sign Language)

Proefschrift ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. dr. J.H.J.M. van Krieken,
volgens besluit van het college voor promoties
in het openbaar te verdedigen op

maandag 21 november 2022
om 16.30 uur precies

door

Anita Slonimska
geboren op 1 mei 1990
te Riga (Letland)

Promotor:
Prof. dr. Asli Özyürek

Copromotor:
Dr. Olga Capirci (Istituto di Scienze e Tecnologie della Cognizione, Italië)

Manuscriptcommissie:
Prof. dr. Onno Crasborn
Prof. dr. Edward Gibson (Massachusetts Institute of Technology, Verenigde Staten)
Prof. dr. Pamela Perniss (Universität zu Köln, Duitsland)
Dr. Erin Wilkinson (The University of New Mexico, Verenigde Staten)
Dr. Sabina Fontana (Universitá di Catania, Italië)

The work described in this thesis was carried out at the Institute of Cognitive Sciences and Technologies, National Research Council of Rome, Italy, Centre for Language Studies, Radboud University and Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands, with financial support from the DCOMM Grant, the European Union's Horizon 2020 research and innovation program under the Marie Sklodowska-Curie Actions Grant Agreement No. 676063.

Mammai un tētim

Table of contents

Chapter 1

General introduction

Linguistic expression is not restricted to the spoken modality. Language can be also expressed and perceived through the visual modality, that is, by the visible articulators of the body, as is the case in sign languages, the natural languages of deaf communities[1]. Research in the last decades has shown that sign languages share the same fundamental linguistic properties as spoken languages at all levels of linguistic organization (Klima & Bellugi, 1979; Liddell, 1980; Padden, 1988; Sandler & Lillo-Martin, 2006; Stokoe, 1960; Wilbur, 1987). Namely, linguistic structures can show similar patterns in both modalities, and it has been shown that similar brain areas support the processing of both spoken and sign languages (e.g., Emmorey, 2001; Emmorey & Özyürek, 2014). However, the visual modality can also give rise to modality-specific linguistic forms and structures, wherein sign languages, unlike speech, can express multiple semantic information units iconically and simultaneously. Recent research has shown that the iconic and simultaneous properties of sign languages constitute a more central role than previously thought in how they are structured (Cuxac, 2000; Padden et al., 2015, 2013; Vermeerbergen et al., 2007), as well as how they are processed and learned and how they develop (for a review, see Perniss & Vigliocco, 2014; Vigliocco et al., 2014). However, the role such properties play in communicative efficiency—a fundamental feature that is known to shape language structure in spoken languages (Gibson et al., 2019; Levshina, *in press*; Levshina & Moran, 2021)—has not been systematically explored in sign languages. To fill this gap, this thesis investigates: a) whether and how the modality-specific properties of simultaneity and iconicity play a role in efficient communication in sign languages; and b) whether simultaneity and iconicity constitute emergent linguistic properties of sign languages that allow for communicative efficiency, rather than an outcome of the general affordances of the visual modality.

One of the fundamental functions of language is to allow efficient communication, that is, "easy, fast and robust information transmission" (Gibson et

[1] The role of the visual modality in linguistic expression also manifests itself in spoken languages, such as in co-speech gestures, which highlights the intrinsically multimodal nature of language (e.g., Clark, 1996; Kendon, 2004, 2014; Özyürek, 2021; Ozyurek & Woll, 2019; Perniss, 2018; Vigliocco et al., 2014). The contributions of co-speech gestures to communicative efficiency are beyond the scope of the present thesis but are discussed where relevant.

al., 2019, p. 389), thereby minimizing costs for producer and perceiver (Gibson et al., 2019; Grice, 1975; Levshina & Moran, 2021). Recent research has argued that this pressure might explain several aspects of language structure crosslinguistically, at least for spoken languages (for a review, Gibson et al., 2019; Levshina, *in press*; Levshina & Moran, 2021), which are fully constrained by the *linearization problem* (Levelt, 1980, 1981). This problem implies that language users are required to decompose a holistically perceived event into smaller elements and organize them on a linear scale for linguistic encoding. Given the fleeting nature of the sequentially unfolding linguistic signal and the limits of human memory capacity, the need to linearize pressures languages to be optimized for communicative efficiency (Christiansen & Chater, 2016; Gibson et al., 2019; Jiang & Liu, 2015; Levshina & Moran, 2021; Liu, 2008). Given this pressure, one of the ways that languages ensure communicative efficiency is to arrange syntactically and semantically related elements as close together as possible in a linguistic signal for faster representation access (Bybee, 1985, 2013; Futrell et al., 2015; Gibson, 1998, 2000; Gibson et al., 2019; Gildea & Temperley, 2010; Grodner & Gibson, 2005; Hawkins, 2004; Liu et al., 2017; Temperley, 2007; Temperley & Gildea, 2018). This principle has been shown for spoken languages but not for sign languages, where the visual modality's distinct affordances might influence how this principle is operationalized.

The ability in sign languages to organize information simultaneously and iconically might allow for semantically related elements to be arranged closer together not only sequentially but also simultaneously. That is, the elements of an event that are simultaneously perceived in the world (e.g., a person holding a dog in their hand and petting it at the same time) can likewise be simultaneously represented in sign languages by means of the use of multiple articulators and linguistic strategies that rely on motivated form-meaning mappings—i.e., iconicity. Thus, the nature of the linearization problem and the strategies used to achieve communicative efficiency might be partly dependent on the modality of language and the affordances the modality provides for linguistic expression. For sign languages, there has been no systematic exploration of whether and how simultaneous and iconic constructions are recruited to cluster related meanings closer together for efficient communication or which linguistic strategies (i.e. *lexical signs*, *depicting constructions* and/or *constructed action*; these will be discussed in detail in Section 1.1.1.1.) are used to achieve communicative efficiency.

Furthermore, it is not known how simultaneous and iconic constructions emerge in a language. It has been argued that languages, including sign languages, evolve from holistic and iconic representations into systems that become more linearly segmented and arbitrary, akin to structures in speech (Aronoff et al., 2005; Fay, Ellison, et al., 2014; Fay et al., 2013; Garrod et al., 2007; Goldin-Meadow et al., 2008; Motamedi et al., 2019; Özyürek et al., 2015; Senghas, 2019; Senghas et al., 2004, 2010, 2013; Theisen et al., 2010). The argument holds that such segmented elements can then allow compositionality, i.e., the recombination of meaning units in systematic ways to construct new meanings (Kirby et al., 2015; Motamedi et al., 2019; Özyürek et al., 2015; Smith et al., 2003, 2013), which arises due to languages becoming optimized for communicative efficiency and language learning (Kirby et al., 2008, 2015; Motamedi et al., 2019). However, in this respect, the abundant use of simultaneous and iconic constructions in conventional sign languages presents a peculiar phenomenon that requires systematic inquiry with regard to their function. Considering that the emergence of compositional structure has been predominantly studied through patterns of linearization, we do not know whether sign languages take advantage of the affordances of the visual modality to devise *simultaneous compositionality* due to the pressures of communicative efficiency. The use of simultaneous and iconic constructions might constitute a linguistic property of sign languages that has emerged as an adaptation of communicative efficiency in the visual modality. To date, no research has investigated the possible trajectory of simultaneous and iconic constructions in language evolution in relation to their function in communicative efficiency.

Thus, the goal of this thesis is to first explore whether simultaneous and iconic constructions are employed for communicative efficiency. Secondly, if this is found to be the case, the aim is to explore whether these constructions are afforded by the visual modality alone or whether they have potentially emerged in the linguistic system as an adaptation for communicative efficiency. Such an inquiry would provide a novel contribution to our understanding about the function of simultaneity and iconicity in sign languages and further inform the field of communicative efficiency by illuminating general and modality-specific strategies that language users adopt to achieve it. This will also contribute to the field of language evolution by taking into account not only the well-explored and -documented emergence of linearization but also the emergence of the simultaneous and iconic organization of meaning elements as a linguistic property.

In this thesis, I focus on the use of simultaneous and iconic constructions by LIS (Italian Sign Language) signers **(Chapters 2 and 3)** and by Italian speakers who were asked to use only gestures to communicate (i.e., as in the so-called *silent gesture* paradigm, see Goldin-Meadow, So, Özyürek, & Mylander, 2008) **(Chapter 4)** to explore whether the visual modality-specific properties are recruited for efficient communication and how they might have emerged. In order to provide evidence for the communicative function of these properties, I will explore them from different but complementary perspectives in terms of their role in: information organization, linguistic encoding strategies, and language evolution. To explore these perspectives, I pose three research questions, which I investigate in the three experimental studies described in this thesis:

(RQ1) Are visual modality-specific properties in the form of simultaneous and iconic constructions recruited for efficient communication in LIS?

(RQ2) What linguistic strategies are used for efficient communication in LIS?

(RQ3) Do simultaneous and iconic constructions constitute an emergent property of sign languages or a mere affordance of the visual modality?

By answering these questions and bringing together the fields of sign languages, communicative efficiency and language evolution[2], I attempt to go beyond the previous literature and provide a novel understanding of whether and how modality-

[2] Considering that the term *language evolution* "has three common interpretations: biological evolution, language change, and the cultural emergence of linguistic structure" (Tamariz & Kirby, 2016, p.37), note that in the present thesis I am focusing on the latter interpretation and use the terms *language evolution* and *language emergence* interchangeably to refer to the cultural emergence of iconic and simultaneous linguistic structures.

specific properties are recruited for communicative efficiency and whether they evolve to be optimized for this function. Furthermore, by assessing these questions by means of an experimental design controlling for the amount and type of information to be encoded, the contribution of this thesis lies not only in new theoretical knowledge but also a novel methodology, thereby laying the groundwork for further cross-modal and cross-linguistic inquiry in different population groups.

1.1. Previous research

In the next sections I explain the necessary theoretical background for the studies described in **Chapters 2–4**. In Section 1.1.1. I outline how iconicity is used for linguistic organization by means of various linguistic strategies (*lexical signs, depicting constructions,* and *constructed action*) and how iconicity enables the creation of simultaneous constructions in sign languages. I then describe how simultaneous constructions are categorized in the existing literature and assess their function in sign languages. In Section 1.1.2. I define communicative efficiency and focus on a principle known as *dependency distance minimization,* which reflects in tendency to cluster related meanings closer together in a linguistic structure and has been amply attested in spoken languages. Later, based on the literature outlined in Section 1.1.1., I describe how the use of simultaneous constructions in sign languages can be compared to dependency distance minimization in spoken languages and used to achieve communicative efficiency. In Section 1.1.3. I turn to research on the selective pressures in language evolution that lead to the emergence of segmentability and compositionality in a linguistic system and discuss the potential implications of this research for the function of simultaneous and iconic constructions in sign languages.

1.1.1. Iconicity and simultaneity as structuring properties in sign languages

In this section, I elaborate on what is known about how iconicity and simultaneity are recruited for linguistic organization in sign languages. In the first part of this section

(1.1.1.1.), I focus on two types of iconicity, *imagistic* and *diagrammatic*, and how they are employed via the different linguistic strategies that signers have at their disposal for linguistic encoding on all levels of linguistic organization. I first outline how, on the lexical level, both types of iconicity are employed for structuring *lexical signs*. I then go beyond the lexical level and focus on how imagistic and diagrammatic iconicity are employed in depicting strategies in order to depict events, i.e., showing what they look like (Clark, 2016, 2019), rather than describing events by arranging lexical signs in a linear fashion. In particular, I describe two depicting strategies available to signers: *depicting construction (DC)*, which is used to depict events from the observer perspective, and *constructed action (CA)*, which is used to depict events from the character perspective. I focus on how both depicting strategies allow signers to represent multiple semantic elements of perceptually simultaneous aspects of an event.

In the second part of this section (1.1.1.2.), I outline how simultaneous constructions have been categorized in the literature, define simultaneous constructions that is the focus of the present thesis, and describe the functions that have been attributed to simultaneous constructions in the literature.

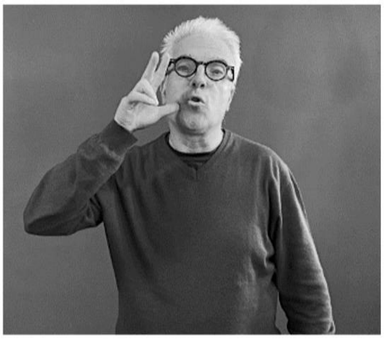
1.1.1.1. The role of iconicity in linguistic organization in sign languages

Iconicity in language, generally defined as the "existence of a structure-preserving mapping between mental models of linguistic form and meaning" (Taub, 2001b, p. 23), is omnipresent in sign languages on all levels of linguistic organization (Demey et al., 2008; Occhino, 2017; Perniss et al., 2010; Taub, 2001). Two types of iconicity can be differentiated in sign languages—*imagistic* and *diagrammatic* (Taub, 2001). While imagistic iconicity refers to the resemblance between the form of the sign and its meaning, diagrammatic iconicity manifests in the motivated relationship between components, i.e., the relationships between phonological parameters in single signs, or the relationships between individual signs in the signing space (for an overview of conflicting views on imagistic and diagrammatic iconicity in the literature, see Perniss, 2007a). These two types of iconicity are not mutually exclusive and they interact when encoding meaning at many levels of linguistic structure—phonology, morphology, lexicon and syntax (Meir et al., 2013; Occhino, 2017; Perniss et al., 2010), often blurring the line between these domains (Lepic & Occhino, 2018; Russo, 2004; Wilcox, 2004).

1.1.1.1.1. Iconicity on the lexical level: lexical signs

Lexical signs, also called *lexical units (LU)*, constitute conventional linguistic forms for concepts and are comparable to words in spoken languages (Johnston & Schembri, 1999). A lexical sign is constructed from manual sign parameters, i.e., *handshape, location, movement,* and *hand orientation,* and it can also include non-manual sign parameters, i.e., *torso, mouth, eye gaze* and *facial expression* (Volterra et al., 2022). Lexical signs are often accompanied by a *mouthing*, i.e., voiced or unvoiced pronunciation of the corresponding spoken word, or by a *mouth gesture*, i.e., a mouth movement that does not correspond to the spoken word (Boyes-Braem & Sutton-Spence, 2001; Crasborn et al., 2008; Roccaforte & Volterra, 2016; Sutton-Spence, 2007). Sign parameters, which constitute the smallest building blocks of a lexical sign, have long been considered analogous to the smallest building units of spoken languages, i.e., phonemes. The difference between constructing a word and a sign, however, lies in the fact that while phonemes are arranged linearly in spoken languages, in sign languages parameters are arranged simultaneously. Furthermore, some research indicates that sign language parameters, unlike phonemes in spoken languages, are not meaningless units but may have instead inherently iconic properties (Boyes-Braem, 1981; Occhino, 2017) and can be motivated by cognitive perceptual features (Fuks, 2014).

A lexical sign can, but does not always have to, retain a resemblance to its referent. For example, the sign in LIS for *woman* (Figure 1-1a) is considered opaque or arbitrary, implying that its meaning is difficult if not impossible to guess from the form of the sign (Boyes-Braem et al., 2002). On the other hand, for iconic signs, the sign's resemblance to its referent is more evident, e.g., the sign for *bird* (Figure 1-1b), and as such this lexical sign is more transparent in regard to its form-meaning mapping.



a **b**

Figure 1-1. Opaque lexical sign in LIS for *woman* (a) and transparent iconic lexical sign in LIS for *bird* (b).

The sign for *bird* in LIS is performed by a closed fist and an extended thumb and index finger that open and close. This clearly has an imagistic iconic resemblance to the beak of a bird, and the connection is further strengthened by the location of the sign, which is performed in front of the mouth. The same lexical sign is also used in ASL (American Sign Language), while a very different sign for *bird* is used in TİD (Turkish Sign Language). The variations in the way in which a concept can be articulated points to what Taub (2001) refers to as the *analogue-building model*, which is thoroughly described by Emmorey (2014, p.2) by contrasting the structural properties of the sign for *bird* in ASL and TİD (Figure 1-2).

The analogue-building model implies a multi-step process. First, an image representing the concept has to be selected, e.g., the head of the bird in the case of ASL and LIS and the entire body of the bird in the case of TİD. Then, the selected image has to be schematized by identifying the relevant parts of the concept as well as the relationship between these parts, e.g., the bird’s head and beak and the movement of the beak for ASL and LIS vs. the bird’s body and wings and the movement of the wings for TİD. Finally, a linguistic form has to be chosen in order to map the schematized image onto visual forms, e.g., a one-handed sign consisting of a closed fist and extended thumb and index fingers that open and close in front of the mouth of the signer for ASL and LIS vs. a two-handed sign consisting of flat palms facing downwards and moving up and down at the signer’s sides for TİD.

Accordingly, lexical signs are imagistically iconic as a whole, i.e., the form of a sign resembles the meaning (Perniss, 2007a). They can nevertheless be considered diagrammatically iconic with respect to how they have been constructed, i.e., schematization of meaning components mapped onto different sign parameters and their relationship to each other (Emmorey, 2014; Pietrandrea & Russo, 2007).

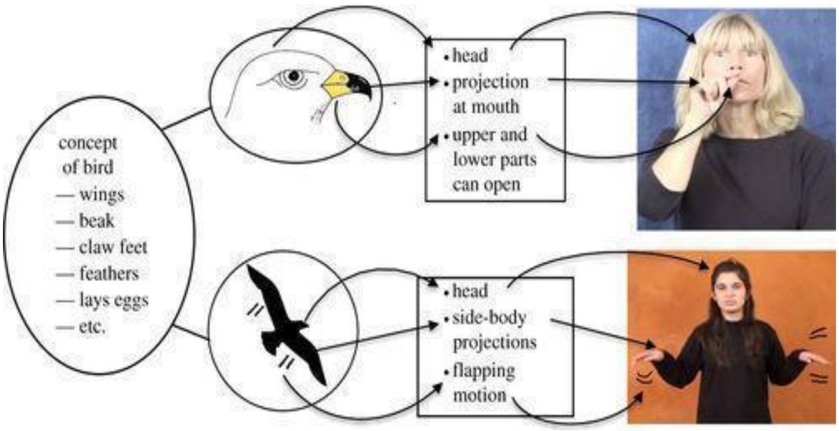


Figure 1-2. An example of the sign for *bird* in ASL (top) and TİD (bottom) based on the analogue-building model. Reprinted from "Iconicity as structure mapping" by Emmorey, K., 2014, *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1651), 20130301, p. 2., (https://doi.org/10.1098/rstb.2013.0301). CCBY.

When it comes to the function of iconicity in sign languages, it has been studied mostly at the lexical level. Previous research has shown that lexical signs with salient iconic properties have a processing advantage during comprehension (Grote & Linz, 2003; Perniss et al., 2010; Perniss & Vigliocco, 2014; Thompson et al., 2009; Vigliocco et al., 2014; Vinson et al., 2015). Furthermore, it has been argued that iconic lexical signs help ground linguistic concepts in real-world representations, and as such, might support initial sign language development (Ortega, 2017; Ortega & Morgan, 2015; Perniss et al., 2010; Perniss & Vigliocco, 2014; Thompson et al., 2012; Tolar et al., 2008).

1.1.1.1.2. Iconicity beyond the lexical level: depicting constructions and constructed action

Iconicity in sign languages manifests not only in labeling single concepts (i.e., lexical signs) but it can be also used for depicting events which are part of a higher-level representation. Signers can *describe* an event by using linear arrangement of the lexical signs, comparable to how individual words are combined to form a sentence in spoken languages. However, signers can also show what the event looks like by *depicting* it through iconic constructions (see Figure 1-3) (Clark, 2016, 2019; Cuxac, 1999, 2000).

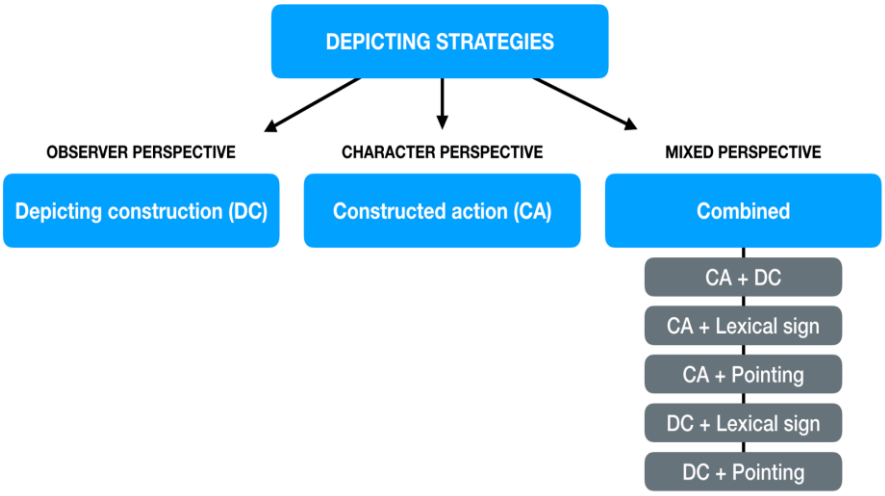


Figure 1-3. Depicting strategies in sign languages from observer and character perspectives and their combinations with other linguistic strategies in sign languages.

Such so-called *depicting strategies* allow speakers to make use of iconicity, multiple body articulators and space to depict multiple meaning elements (*imagistic iconicity*) and their relationship to each other (*diagrammatic iconicity*) simultaneously when encoding an event (Perniss, 2007a; Risler, 2007). Signers have two depicting strategies at their disposal: *depicting construction* (DC) and *constructed action* (CA). Depicting constructions take the vantage point of the observer and accordingly depict events on a miniature scale in the signing space in front of the signer, i.e., *observer perspective* (Figure 1-4) (Perniss, 2007a, 2007b). Constructed actions take the vantage point of a character and accordingly depict events from the viewpoint of a specific referent and on a life-sized scale, i.e., *character perspective* (Figure 1-5). Furthermore, different depicting strategies as well as lexical and pointing signs can be used in combination in a single construction, leading to a depiction from a mixed perspective (Jarque & Pascual, 2016; Perniss, 2007a, 2007b).

Depicting construction (DC): Depicting strategy from the observer perspective

Use of depicting constructions (DC), also known as *classifier constructions*, *classifier predicates, polycomponential verbs, polymorphemic verbs* (for discussion, see Schembri, 2003), allows a signer to adopt the observer's perspective of the event, as if viewing it from afar on a miniature scale. In a depicting construction, the elements of the scene or event are depicted by means of one or both hands in the signing space in front of the signer (Perniss, 2007a, 2007b; Schembri, 2003). A specific handshape and its motion (if present) are directly associated with the size, shape, and motion of the referent through imagistic iconicity (Boyes-Braem, 1981; Brennan, 1990; Perniss, 2007a). For example, in LIS a flat handshape is used to represent the flat, solid surfaces of objects, e.g., a book (see Fig. 1-4), a table, a shelf, a page, a wall, etc., while an extended index finger is used to represent thin and elongated objects, e.g., a pen (see Fig. 1-4), a toothbrush, a key, a strap on a bag, a person, etc. Such depicting handshapes are used to locate objects in the signing space from the observer perspective. Handshapes can be altered to encode information about the location, orientation and movement of the object in space. For example, the orientation of an index finger depicting a pen can be changed to match the position of the pen's location in the real world. In other words, depicting constructions are both imagistically and diagrammatically iconic—they establish semantic relationships between the objects and space or other objects (Emmorey, 2002; Liddell, 2003; Sümer et al., 2013). For instance, after introducing the referents *book* and *pen* by means of lexical signs, a signer can use a depicting construction to encode the referents and their spatial relationship to each other, i.e., the pen positioned to the right of the book (Figure 1-4).

RH: lexical sign: *book* lexical sign: *pen* DC: *pen*
LH: lexical sign: *book* DC: *book*

Figure 1-4. A signer depicting a pen positioned to the right with respect to a book by means of depicting construction in LIS (final frame). Colored circles represent different referents (blue - pen, red - book).

Thus, depicting handshapes are representative of imagistic iconicity in relation to their respective referents, i.e., a depicting handshape resembles its referent. However, the entire depicting construction is based on diagrammatic iconicity, i.e., the relationship between depicting handshapes reflects the relationship between referents in the event (Perniss, 2007a). Note that such encoding provides direct access to real-world representation, i.e., both objects and their relationship to each other can be represented simultaneously. As such, depicting constructions are suitable for encoding spatial relationships between whole entities e.g., the pen next to a book (as in Figure 1-4), two people standing next to each other, or alternatively with intransitive actions, e.g., a ball rolling, two people approaching each other (Engberg-Pedersen, 1993; Kimmelman et al., 2019; McDonald, 1982; Özyürek & Perniss, 2011; Sümer, 2015; Zwitserlood, 2003).

Recent research has shown that use of the iconicity present in the linguistic organization of sign languages results in enhanced visual attention to spatial relationships (e.g., the pen is to the right of the book) and thus can influence how spatial relationships are conceptualized during message preparation (Manhardt et al., 2020). Furthermore, research investigating the encoding of spatial relationships has shown that signers prefer to depict these relations through depicting constructions rather than using the lexical signs such as *left/right* (ASL: Emmorey, 2002; DGS:

Perniss, 2007a; TİD: Sümer, 2015; Karadöller et al., 2020; NGT: Manhardt et al., 2020). Finally, it has been proposed that this preference might be driven by signers taking advantage of the iconic properties of depicting constructions, which allow spatial information to be encoded in a more informative way in comparison to using solely lexical signs (Manhardt et al., 2020).

Constructed action (CA): Depicting strategy from the character perspective

The visual modality also allows signers to adopt the depicting strategy known as *constructed action* (Metzger, 1995; Tannen, 1989). Constructed action[3] (CA) lets the signer take a character's perspective by depicting an event through the use of iconic mapping, where the articulators of the signer's entire upper body (hands, torso, head, mouth, facial expression, eye gaze) are directly mapped onto the body of the referent(s) and is suitable for encoding animate referents, their actions and their mental states (Cormier, Fenlon, et al., 2015; Cormier, Smith, et al., 2015; Kurz et al., 2019; Metzger, 1995; Quinto-Pozos, 2007a, 2007b). For example, in order to depict that a man is painting and looking at the wall in front of him (Figure 1-5), the signer introduces the referents, i.e., the wall and the man, by means of lexical signs, and then maps the animate referent, i.e., the man, onto their own body. The signer depicts the action of holding a paintbrush with the right hand, sweeping it in an up-and-down motion in the signing space while also employing the torso, head, facial expression, and eye gaze to depict the man.

While *constructed action* is the most frequently used term, it is also known as *transfer of person* (Cuxac, 1999, 2000, 2001; Fusellier-Souza, 2006; Russo, 2004), *role shift* (Padden, 1986; Quer, 2018), *referential shift* (Emmorey, 2002; Engberg-Pedersen, 1993), *body classifier* (Supalla, 1982, 1986, 2003), and *point of view predicate* (Lillo-Martin, 1995). Given its nature, CA is particularly useful for clearly

[3] Note that in the present thesis CA includes handshapes depicting manipulation of the object. These have been considered elsewhere to be a specific type of depicting construction/classifier construction, i.e., *handling*. I chose to include handling in CA by following Cormier, Smith, et al. (2015), who stress that it is problematic to separate out when handling is produced as a depicting construction or as a direct depiction of the event through constructed action. As a result, I adhere to Cormier, Smith, et al.'s conclusion that "CA and representations of handling/manipulation are really the same phenomenon – i.e., hands of the signer represent hands of the referent" (2015, p. 184).

communicating what a referent does and/or feels and how it interacts with the world by depicting it more truthfully instead of drifting away into a more abstract description, i.e., categorical representation through linearization of lexical signs (Clark, 2016, 2019; Clark & Gerrig, 1990; Quinto-Pozos, 2007a, 2007b). CA can be particularly useful when encoding actions by and interactions between animate referents, considering that a signer can depict the event from the perspective of the character, as in the example described above. Thus, CA allows the agent of an action to be fully present in the utterance since the signer is not merely describing meaning but showing it (Cuxac, 1999, 2000; Ferrara & Halvorsen, 2017), making two semantic elements available to the interlocutor at the same time—the agent and the action (e.g., the man and the action of painting).

RH:	lexical sign: *wall*	lexical sign: *man*	CA: *painting a wall*
LH:	lexical sign: *wall*		
Torso:			CA: *man*
Head:			CA: *man*
Face:			CA: *man*
Eye gaze:			CA: *man*

Figure 1-5. A signer depicting a man painting a wall by means of CA (final frame). Circles represent semantic information units for the same referent (the man and the man's action of painting). White dashed lines represent movement of the hand/s.

Furthermore, as the above example illustrates, CA is representative of not only imagistic iconicity (i.e., a specific part of the signer's body and the action performed by this body part resemble the corresponding body part and action of the referent) but also diagrammatic iconicity, which reflects an iconic relationship between the articulators employed and the relationship between the meaning elements they depict, similar to what is done in depicting constructions (see Fig. 1-4). Thus, CA is not purely imagistic in terms of it being a holistic iconic representation, as has been claimed by some researchers in previous studies (Ferrara & Johnston, 2014; Hodge & Ferrara, 2014; Jantunen, 2017), but it also establishes diagrammatically iconic relations between the elements of the event that can also be encoded by means of different linguistic strategies such as lexical signs, depicting constructions, and pointing (for more details, see **Chapter 3**).

Constructed action is considered to be a strategy used mostly for narrative discourse (Ferrara, 2012) and thus it has been studied almost exclusively in narrative contexts (Cormier, Smith, et al., 2015; Hodge et al., 2019; Hodge & Johnston, 2014; Jantunen, 2017; Pizzuto et al., 2006, among others). The function of CA in narratives has been attributed to both a referential function that expresses core arguments and ensures cohesion (Cormier, Smith, & Zwets, 2013; Ferrara & Johnston, 2014; Hodge & Ferrara, 2014; Hodge & Johnston, 2014; Jantunen, 2017; Pizzuto et al., 2006) and an evaluative function that enhances the narration by making it visually more expressive and entertaining (Dudis, 2002; Mather & Winston, 1998; Poulin & Miller, 1995; Roy, 1989; Wilson, 1996; Winston, 1992). Considering that the referential function of CA has not been assessed outside of narrative context, it is impossible to tease apart whether CA is used for referential function only, evaluative function only, or the combination of the two. Arguably, out of all the linguistic strategies at signers' disposal, CA as a fully linguistic device has received the least attention and the referential function of CA as well as its role in efficient communication remains little understood.

Importantly, the "linguistic status" of depicting in contrast to describing has been debated and depicting is mainly considered to be gestural and thus non-linguistic or, at best, partly linguistic (see Goldin-Meadow & Brentari, 2017). In such a view, the use of CA is similar to gesture in spoken languages, i.e., to provide additional information in an analogue and gradient manner alongside the linguistic signal, which is characterized by its categorical and discrete properties. However, recent views in cognitive linguistics invite reconsideration of this notion and propose that language should instead be viewed as comprising composite utterances of different semiotic levels (e.g., Bybee, 2010; Cienki, 2016; Enfield, 2009; Ferrara & Halvorsen, 2017;

Hodge & Ferrara, 2014; Jantunen, 2017) and that *depicting* can be considered to be as linguistic as *describing* is (Capirci et al., 2022; Cuxac, 2000; Garcia & Sallandre, 2020; Kurz et al., 2019). In the present thesis I adopt this broader view of language and aim to provide further evidence for it by showing that properties specific to the visual modality in CA can be recruited to fulfill core linguistic functions such as communicative efficiency.

1.1.1.2. The role of simultaneity in linguistic organization in sign languages

In sign languages, simultaneity can also be present at the lexical level and beyond. At the lexical level, simultaneity is used to organize sign parameters in order to create a lexical sign. Beyond the lexical level, simultaneity is used to organize higher-level structures through simultaneous constructions, which are defined as "representations that are produced in more than one articulatory channel, whereby each channel bears distinct and independent meaning units which stand in some relationship to each other" (Perniss, 2007a, p. 39).

Simultaneous constructions can be categorized according to their form and linguistic function. Below I outline how simultaneity has been categorized in the literature and describe what type of simultaneity is the focus of the present thesis.

The one and only volume devoted to simultaneity in sign languages (Vermeerbergen et al., 2007) differentiates simultaneity based on three forms: manual simultaneity, manual-oral simultaneity, and simultaneous use of other (manual and non-manual) articulators (p. 2).

Manual simultaneity refers to simultaneous constructions in which each hand encodes different information. It is possible for both hands to sign different meanings simultaneously or, alternatively, one of the signs can be held constant while the other hand continues signing. The latter example is often connected to discourse when listing items, preserving topic and focus or using depicting constructions to encode relative locations.

Manual-oral simultaneity refers to the simultaneous use of hands and mouth. Research on this kind of simultaneity has concentrated on the information expressed when signs are combined with *mouthings,* i.e., articulation of words from a spoken language, or with *mouth gestures* which provide additional information about the event being encoded by the hands.

Simultaneous use of other (manual and non-manual) articulators not only refers to simultaneity that is achieved by means of both hands or hands and mouth, it also includes, as the name suggests, other non-manual articulators such as the signer's torso, head, eye gaze, and facial expression. Research on this form of simultaneity studies the simultaneous encoding of different perspectives of the same event and combinations of manual signs and lexical non-manual items (e.g., question words encoded by non-manual articulators).

Perniss (2007a) takes a different approach to categorizing simultaneous constructions. Instead of categorizing them based on the form, Perniss looks at linguistic function and subsumes simultaneous constructions under two categories: simultaneity of perceptual structure for encoding spatial/temporal organization, and simultaneity of discourse structure for organizing discourse. Points 1 to 3 below are representative of the perceptual structure, while points 4 to 6 are representative of the discourse structure:

Functions of simultaneous constructions:

1. referent representation on both hands to express spatial information (in the depictions of the spatial relationship between referents)
2. referent representation on both hands to express temporal and spatial simultaneity of events (in the depiction of action or interaction between referents)
3. the expression of temporal simultaneity of events or states (aspectual information)
4. the hold of a topic on one hand while the other hand signs related information (topic-comment structure)
5. the hold of an enumeration morpheme on one hand while the other hand signs one or more related signs
6. the hold of index sign on one hand while the other hand signs one or more related signs (Perniss, 2007a, p. 40)

Research on the communicative rather than purely linguistic function of simultaneous constructions is still in its infancy. Perniss (2007a), one of the few researchers who has highlighted the role of iconicity in creating simultaneous constructions, has proposed that simultaneous and iconic constructions can be used to increase informativeness and efficiency in narrative discourse. In her research, Perniss focuses on simultaneous constructions that encode different perspectives of the same event, known as *non-aligned character and double-perspective constructions*, which involve the mixing of constructed action and depicting constructions. She argues that use of such a strategy is both informative and efficient in that by using it "signers can achieve full semantic specification of a spatial event" (Perniss, 2007a, p. 219) and that it "ensures a high degree of informativeness, encoding multiple event components, and contributing to overall spatial coherence" (p. 220). This research suggests that combining depicting strategies from character (CA) and observer (DC) perspectives may function to increase informativeness and efficiency in narrative discourse. Some descriptive research has also shown that signers can vary the amount of information that can be simultaneously encoded in a sign language (Dudis, 2004; Napoli & Sutton-Spence, 2010; Risler, 2007) and thus they can use simultaneity as a tool to package more information simultaneously when necessary (for more details, see **Chapter 2**).

Taken together, these few insights indicate that simultaneous and iconic constructions might potentially be used for achieving communicative efficiency. However, whether simultaneous and iconic constructions are actually recruited outside of narrative contexts and serve to achieve communicative efficiency remains unexplored.

In this thesis, I am focusing on simultaneous and iconic constructions used to encode perceptually simultaneous elements of an event. With regard to form, I am concerned with the simultaneous use of manual and non-manual articulators. Thus, I am focusing on the use of multiple articulators for the simultaneous encoding of distinct semantic elements pertaining to the same perceptually simultaneous event. More specifically, I am asking whether LIS signers employ such simultaneous and iconic constructions in order to achieve communicative efficiency in an experimentally controlled interactive setting.

Summary

In this section (1.1.1.), I outlined how imagistic and diagrammatic iconicity can be recruited to structure sign languages at the lexical level and beyond. I described how iconicity allows information to be structured in a simultaneous and iconic manner, both through the simultaneous combining of parameters to create a single sign and through the simultaneous encoding of complex events by combining multiple articulators and different linguistic strategies. A crucial point is that unlike the strategy of using only lexical signs and concatenating them one by one to express a complex event, depicting strategies such as depicting construction and constructed action and their potential to combine with lexical signs mean that signers are not required to split and arrange each piece of information on a linear scale. Instead these constructions allow signers to express perceptually simultaneous events through simultaneous constructions. I then described how simultaneous constructions have been categorized in the literature and outlined the relatively limited existing research on their functions. Finally, I noted that in the present thesis I am focusing on simultaneous and iconic constructions that employ manual and non-manual articulators to depict multiple semantic elements of perceptually simultaneous events and whether such constructions are used to increase communicative efficiency, which I elaborate in more detail in the next section.

1.1.2. Efficient communication in spoken and sign languages

In the previous section I described how the visual modality-specific properties of iconicity and simultaneity allow multiple semantic elements pertaining to the same event to be encoded not only in a linear fashion, as they would be in spoken languages, but also simultaneously. I also noted that some research indicates that these properties might potentially be used for communicative efficiency. However, there is no systematic research on communicative efficiency in sign languages. Therefore, in this section I assess how communicative efficiency has been studied in spoken languages and draw parallels with how it could function for sign languages and the role that the visual modality-specific properties might play in achieving it.

Given that language unfolds in the temporal dimension in both spoken and sign languages, it is constrained by what Levelt defines as the *linearization problem* (1980, 1981), which implies that the mapping of thought onto language must be expressed in a linear order. That is, pieces of information in the form of sounds, words or signs are replaced by new pieces of information. There are two main considerations to take into account in this regard: the loss of signal and the rate of incoming linguistic input (Christiansen & Chater, 2016). It is estimated that the acoustic signal is lost after 50-100ms (Elliott, 1962; Remez et al., 2010) and maintaining visual information is barely possible after 60-70ms (Pashler, 1988; Sperling, 1960). Speakers of spoken languages, at a normal speech rate, produce on average 5-6 syllables per second (Studdert-Kennedy, 1986). The production of signs, although slower than sounds, are also fast visual events, i.e., an average of a quarter of a second per syllable (Wilbur & Nolen, 1986). Working memory is also limited for both auditory and visual stimuli (Boutla et al., 2004; Cowan, 2001, 2010; Daneman & Green, 1986; Miller, 1956; Wilson & Emmorey, 2006; but see Rudner & Rönnberg, 2008 for a review on similarities and differences in working memory for languages of different modalities). This means that the incoming signal must be processed rapidly, before it fades or is overwritten by new input (*now-or-never-bottleneck*, Christiansen & Chater, 2016).

The fast and fleeting nature of the linguistic signal pressures languages to be communicatively efficient (Christiansen & Chater, 2016; Gibson et al., 2019). Communication can be considered efficient if language "is structured to facilitate easy, rapid, and robust communication" (Gibson et al., 2019, p. 389). It is crucial to point out that as communication involves two parties, communicative efficiency involves minimizing effort not only for the producer but also for the comprehender to ensure that information can be decoded correctly (Gibson et al., 2019; Grice, 1975). Research on spoken languages has provided a plethora of evidence showing that linguistic structure is indeed optimized for efficient communication on all levels of linguistic organization (Aylett & Turk, 2004; Futrell et al., 2020; Gibson et al., 2013, 2019; Levshina, *in press*; Levshina & Moran, 2021; Pellegrino et al., 2011; Piantadosi et al., 2011; Zipf, 1949). In the present thesis, I focus on the achievement of communicative efficiency at higher-level structures by means of organizing syntactically and semantically related meanings as close together as possible, a universal tendency of languages (Gibson, 1998, 2000; Gibson et al., 2019; Hawkins, 2004; Levshina & Moran, 2021).

It has been established that in spoken languages words that are dependent on each other for interpretation are encoded as close to each other as possible in the spoken signal. This strategy has been termed *dependency distance/length minimization* (Futrell et al., 2015; Gildea & Temperley, 2010; Grodner & Gibson, 2005; Hawkins, 2004; Jaeger & Tily, 2011; Liu et al., 2017; Temperley, 2007) but it is also known as the *dependency locality* (Gibson, 2000, 2020) and the principle of *minimize domains* (Hawkins, 2004, 2014). This strategy is considered to be driven by languages being structured "to enable maximal information transfer with minimal effort" (Futrell et al., 2020, p. 374). Specifically, assuming that production and comprehension proceed incrementally, integrating new words into the structure requires previous material be retrieved (Gibson, 1998, 2000). For example, the sentence *John threw the old trash sitting in the kitchen out* is more costly to process than *John threw out the old trash sitting in the kitchen*, as there is a longer distance between the words (*threw* & *out*) that are dependent on each other for interpretation of the sentence (Futrell et al., 2015, p. 10337). Accordingly, longer distances between related words lead to greater processing difficulty (Futrell et al., 2020; Gibson, 1998; Hawkins, 2004, 2014; Jaeger & Tily, 2011).

Minimizing dependency distances promotes efficiency in communication, as it leads to lower cognitive load on working memory in both production and comprehension relative to longer dependencies, which take longer to process (Ferrer i Cancho, 2004; Gibson, 1998, 2000, 2000; Hawkins, 2004, 2014; Jiang & Liu, 2015; Liu, 2008). As a result, efficient communication can be achieved through faster representation access when related meanings are located closer together (Gibson et al., 2019; Hawkins, 2004). Support for this assumption comes from corpus studies (see Levshina & Moran, 2021; Liu et al., 2017; Temperley & Gildea, 2018, for a review), as well as experimental studies on language production and comprehension. For example, experimental studies on production show a preference for linguistic structures that lead to shorter dependencies (Fedzechkina et al., 2017; Yamashita &

Chang, 2001) and studies on comprehension show greater ease in processing shorter dependencies (e.g., Bartek et al., 2011; Chen et al., 2008; Fedorenko et al., 2013; Gibson, 1998; Gibson & Wu, 2013; Grodner & Gibson, 2005; Hsiao & Gibson, 2003; Levy et al., 2013).

While dependency distances have not been studied in sign languages, in this thesis I propose that simultaneous and iconic constructions might prove to be useful for achieving communicative efficiency in a way that is comparable to dependency distance minimization in spoken languages, i.e., clustering related meanings as close together as possible. In sign languages related meanings can be clustered together not only by positioning related signs next to each other but also by producing related signs simultaneously, resulting in imagistically and diagrammatically iconic constructions. As we saw in the previous section, research by Perniss (2007a, 2007b) suggests that simultaneous constructions constitute an informative and efficient strategy in narrative discourse. We also know that information density of simultaneity, i.e., the amount of information encoded simultaneously, can be manipulated by signers (Dudis, 2004; Napoli & Sutton-Spence, 2010; Risler, 2007). However, we are still in the dark with regard to what influences the use of simultaneous constructions and what linguistic strategies are used to express events of varying information density. Furthermore, it is not clear whether simultaneous and iconic constructions can be used outside of narrative discourse and whether they can function to achieve communicative efficiency. By focusing on the general tendency of clustering related meanings closer together, as attested in spoken languages, and assessing it in sign languages in the form of simultaneous and iconic constructions, I aim to test whether the properties of the visual modality that sign languages afford play a role in how communicative efficiency is achieved. Focusing on sign languages can provide novel insights into the general and modality-specific strategies language users adopt to achieve communicative efficiency.

1.1.3. Language evolution and implications for the function of simultaneous and iconic constructions in sign languages

Another issue regarding the function of simultaneous and iconic constructions concerns their role in language evolution. Are they available to signers as general affordances of the visual modality and present from the initial stages of language emergence, or have they evolved as an adaptation to the pressure of communicative efficiency? In the following section, I first outline the literature on the selective pressures in language evolution that lead to the emergence of compositionality, which is possible through the segmentability of holistic representations in communicative systems. I later discuss the implications of this research for the function of simultaneous and iconic constructions in sign languages.

It has been argued that holistic and iconic representations can serve as a base for the emergence of communicative systems, while linearly segmented and arbitrary forms and compositional structure only emerge later, as these systems evolve (Fay, Ellison, et al., 2014; Fay et al., 2013; Fay, Lister, et al., 2014; Garrod et al., 2007; Goldin-Meadow et al., 1996; Kirby et al., 2008; Motamedi et al., 2019; Senghas et al., 2004). Unlike holistic expressions which represent meaning as a single, undividable whole, compositional structure implies that "the meaning of a signal is a function of the meaning of its parts and the way they are put together" (Smith et al., 2003, p. 372). That is, in a compositional system, larger meaning units are systematically constructed from smaller meaning units (e.g., a sentence is constructed of words in spoken and signs in sign languages). As such, compositionality allows the meaning of previously unencountered utterances to be deduced from knowledge of the meaning elements and the ways they can be combined (Smith et al., 2003).

Experimental and computer simulation research provides evidence that compositional structure in language emerges from holistic representations as an adaptation to the selective pressures of language use and language transmission to new learners, which pressure languages to become more communicatively efficient and systematic (e.g., Beckner et al., 2017; Kirby, 2000; Kirby et al., 2008, 2014, 2015; Motamedi et al., 2019; Nölle et al., 2018; Raviv et al., 2019; Theisen et al., 2010). While language use is argued to result in the selection of those linguistic forms that are expressive but easier to use (communicative efficiency), language transmission is argued to result in the selection of those linguistic structures that are compressible and as such easier to learn (i.e., systematicity) (Kirby et al., 2015; Nölle et al., 2018). Some research suggests that only the combination of both pressures, efficient language use during interaction and ease of transmission to new learners, leads to compositional structures that are both informative and efficient (Kirby et al., 2015; Motamedi et al., 2019; Theisen-White et al., 2011). Note that there is other research that suggests that compositionality can also emerge without generational transmission (Nölle et al., 2018; Ortega & Özyürek, 2020; Raviv et al., 2019).

In these accounts, for compositional structure to emerge, holistic representations first have to be segmented into meaning elements. Indeed, segmentability has been recognized as one of the indicators of emergence of linguistic structure, as noted by Goldin-Meadow (2015, p. 174): "Segmentation seems to begin when language begins, although it is not fully developed." Once holistic representations have been segmented into meaning elements, they can be conventionalized through use within the community and re-used systematically for the construction of new meanings, leading to compositionality.

The transition of holistic representations into linearly segmented meaning elements has been documented in naturally evolving linguistic systems in the visual modality, known as *emerging sign languages* and *homesign* systems (e.g., Özyürek et al., 2015; Senghas et al., 2004, 2010, 2013), and with regard to their comparison to *silent gesture* (e.g., Özyürek et al., 2015). The term *emerging sign languages* refers to sign languages used within relatively recently formed deaf communities but that have already transmitted their language to new generations, e.g., Al-Sayyid Bedouin Sign Language in Israel (Aronoff et al., 2008; Sandler et al., 2014), Kenyan Sign Language (Morgan, 2017; Morgan, & Mayberry, 2012), Nicaraguan Sign Language (Kegl et al., 1999; Senghas et al., 2004; Senghas & Coppola, 2001), and Sao Tome and Principe Sign Language (Mineiro et al., 2017). *Homesign* is a term used to refer to gestural systems created by deaf children to communicate with their hearing relatives and who have no exposure to any conventional sign language model (Begby, 2017; Morford, 2002; Özyürek et al., 2015; Senghas et al., 2013). *Silent gesture* is an experimental paradigm in which hearing participants without knowledge of any sign language are asked to invent gestures to communicate certain concepts and which is argued to reflect systematic cognitive biases regarding how the visual-manual modality can express representations based on shared visual-motor imagery (Gibson et al., 2013; Goldin-Meadow et al., 1996, 2008; Ortega & Özyürek, 2020; Özçalışkan et al., 2016; Schouwstra & de Swart, 2014).

Studies that have looked at different cohorts of users of an emerging sign language (Nicaraguan Sign Language, NSL) and compared them with users of homesign systems and hearing participants using silent gesture have shown how emergent properties of the linguistic system can be detected in a gradual increase in segmentability in motion events (Goldin-Meadow, 2015; Özyürek et al., 2015; Senghas et al., 2004, 2010, 2013). Motion events are perceived holistically with regard to their manner and path, and as such they are readily available to be represented simultaneously in a holistic fashion by a single iconic gesture, e.g., representing a rolling down action by lowering the hand while simultaneously moving the hand in a circular motion. These studies show a clear emergent trajectory of representation of motion events starting from holistic and evolving into gradually more linearly segmented representations, i.e., using two separate signs to encode the manner and path of the motion event, arranged in a linear sequence. Namely, holistic forms have been found to be used the most by silent gesturers, followed by homesigners (Özyürek et al., 2015; Senghas et al., 2010) and then the first-cohort NSL signers (Senghas et al., 2010, 2013). In turn, use of segmented forms has been found to gradually increase as linguistic system matures as seen in productions by signers from later cohorts of NSL (Senghas et al., 2004). These studies suggest that regardless of the modality, the signal tends to become more segmented and linearized as language evolves.

We know that established sign languages employ simultaneous and iconic constructions to encode events, as described in Section 1.1.1. However, the evolutionary trajectory of such constructions for depicting complex events has not received any particular attention. To my knowledge, only two studies have addressed the emergence of simultaneous constructions for event encoding, albeit with a slightly different focus than the one taken in this thesis. Senghas & Littman (2004) focus on the emergence of simultaneity in the encoding of motion events by comparing NSL and the established sign language LSE (Spanish Sign Language). They argue that the emergent trajectory of motion event encoding suggests that, relative to the first cohort of NSL signers, simultaneous forms used by LSE signers are less representative of holistic representations. Rather, they are a result of bringing together already segmented elements in a simultaneous manner within a single articulator, i.e., a hand (Senghas & Littman, 2004). Kocab et al. (2016) show that as the language evolves, NSL signers are more likely to use simultaneous constructions where two hands are used to encode overlapping events. The findings of these studies indicate that the emergent trajectory of information organization in a sign language does not stop at segmentation in a linear structure but also allows for the segmented elements of meaning to be arranged in simultaneous constructions (i.e., simultaneous compositionality) as language further evolves (for more details, see **Chapter 4**).

The evolutionary trajectory of the simultaneous constructions achieved through the depicting strategies described above (Section 1.1.1)—which are based on imagistic and diagrammatic iconicity—is not clear. That is, it is not known whether simultaneous and iconic constructions that recruit multiple body articulators and iconicity are available for depicting informationally dense events in sign languages due to the general affordances of the visual modality or whether these affordances have been optimized during language evolution for greater communicative efficiency through simultaneous compositionality. Thus, one of the aims of this thesis is to address this issue, at least partially, by comparing the simultaneous and iconic constructions used by LIS signers and Italian silent gesturers asked to describe events of varying levels of information density.

1.2. The present thesis

In the present thesis, I aim to provide a new perspective on the function of simultaneity and iconicity in achieving communicative efficiency by bringing together the above-described literature on sign languages, communicative efficiency, and language evolution. More specifically, I aim to assess whether the simultaneous and iconic constructions used to encode perceptually simultaneous events are employed to achieve communicative efficiency in an experimentally-controlled interactive context. Such an interdisciplinary approach will not only allow some of the missing gaps outlined in the previous sections of this chapter to be filled, but on a more general scale it will also provide a more thorough understanding of whether and how language is optimized for communicative efficiency based on the affordances provided by the different modalities in which it is realized, and in particular the visual modality.

1.2.1. Research questions

The aim of this thesis is to systematically explore and provide initial insights into the role that modality-specific properties, i.e., simultaneity and iconicity, play in achieving communicative efficiency in sign languages and their emergence. To do so, I investigate three research questions:

(RQ1) Are visual modality-specific properties in the form of simultaneous and iconic constructions recruited for efficient communication in LIS?

(RQ2) What linguistic strategies are used for efficient communication in LIS?

(RQ3) Do simultaneous and iconic constructions constitute an emergent property of sign languages or a mere affordance of the visual modality?

To answer these questions, in the present thesis I use data from adult deaf signers of LIS (Italian Sign Language) and data from adult hearing speakers of Italian in a silent gesture condition. LIS constitutes the conventional sign language used within deaf community in Italy (Volterra et al., 2022*,* see also Box 1-1 for the historical background of LIS).

As described in Section 1.1.1., research on the function of simultaneity and iconicity is very limited, and thus in order to answer *Research Question 1*, in **Chapter 2** I systematically explore whether simultaneous and iconic constructions are recruited for communicative efficiency by LIS signers. My focus on this function is driven by the fact that, as described in Section 1.1.2, research on efficient communication has been limited exclusively to spoken languages (Gibson et al., 2019; Levshina, *in press*; Levshina & Moran, 2021) and thus it is somewhat limited in providing a full understanding of how communicative efficiency shapes language as a general phenomenon and whether modality plays a role in how it is achieved. Thus, focusing on sign languages from the perspective of communicative efficiency is not only beneficial for informing the field of sign language research with respect to the function of the visual modality-specific properties but also for enhancing our understanding of the general and modality-specific strategies that language users adopt to achieve communicative efficiency. As described in Section 1.1.2., in spoken languages the general principle for achieving communicative efficiency takes the form of clustering related meanings encoded by words as close together as possible. In **Chapter 2,** I assess whether signers of LIS take advantage of simultaneous and iconic constructions to achieve communicative efficiency by clustering related meanings closer together not only linearly but also simultaneously. To do so, I use a novel experimental design that is also used in the other two empirical chapters of this thesis (**Chapters 3 and 4**). Furthermore, the same video data is used in **Chapters 2 and 3**.

Once I have gained insights in to how communicative efficiency is achieved with respect to information organization (linear vs. simultaneous), the next logical step is to shift my focus to exploring the specific linguistic strategies used to achieve it, which I assess in **Chapter 3** in order to answer *Research Question 2*. In particular, I focus on the linguistic strategy of *constructed action (CA)*. As described in Section 1.1.1.1., while research has addressed CA with respect to its referential capacity, i.e., to express core arguments and to ensure cohesion (e.g., Cormier, Smith, & Zwets, 2013; Ferrara & Johnston, 2014; Hodge & Ferrara, 2014; Jantunen, 2017; Pizzuto et al., 2006), these studies are limited to narrative discourse. In other communicative contexts, however, our understanding of the role of CA is minimal. I choose to focus on CA for two reasons. First, I aim to assess whether CA is used for communicative efficiency given that its properties of imagistic and diagrammatic iconicity allow for a high degree of flexibility in encoding informatively dense events involving animate referents and their interaction, as described in Section 1.1.1.1. Second, I aim to fill an existing gap in sign language research by providing the first experimentally controlled insights into the use of CA as a referential resource. Accordingly, I aim to go beyond what we already know about CA and the function it fulfills in narrative discourse by assessing the linguistic strategies used for communicative efficiency in a controlled experimental setting and the role that CA plays in achieving it.

The main hypothesis of this thesis is that simultaneity and iconicity in the form of simultaneous and iconic constructions constitute linguistic properties whose function is to achieve communicative efficiency in a sign language such as LIS. If this is the case, it implies that the use of such constructions has emerged and evolved as an adaption to the pressure of communicative efficiency, leading to compositional structure that is not only linear but also simultaneous. Thus, in order to provide more clarity on this topic and answer *Research Question 3*, in **Chapter 4** I aim to assess whether quantitative and qualitative differences in the use of simultaneous and iconic constructions can be detected when comparing their use as part of the linguistic system (i.e., in a sign language) and as a general affordance of the visual modality (i.e., in silent gesture). While research on language evolution has predominantly concentrated on the emergence of compositional structure through linearization (as described in Section 1.1.3.), it is not clear how the simultaneous constructions that signers of established sign languages use might have emerged. Thus, it remains to be explored whether simultaneous and iconic constructions constitute a general affordance of the visual modality and thus can be employed without need of a linguistic system (i.e., silent gesture) or whether they can be considered an emergent linguistic property of a sign language optimized for communicative efficiency through simultaneous compositionality. The final aim of the present thesis is to not only corroborate the hypothesis that simultaneous and iconic constructions are recruited for communicative efficiency from a language evolution perspective but also contribute to the field of language evolution by complementing the existing research on the emergence of linearization with new insights on the emergence of the simultaneous and iconic constructions in a linguistic system.

Considering that an increase in the amount of information that must be communicated leads to greater pressure for communicative efficiency, I manipulate the pressure for communicative efficiency by means of systematically increasing the number of semantic information units that must be communicated in an interactive setting (i.e., *director-matcher* game) through a controlled experimental design, which is used in all three empirical studies of this thesis (**Chapters 2, 3 and 4**). By adopting such an approach, I am able to directly test how pressure for communicative efficiency during interaction is reflected in the simultaneous constructions and linguistic strategies used by signers (**Chapters 2 and 3**) and how the simultaneous constructions used by silent gesturers compare to those of signers (**Chapter 4**). This experimental design will allow me to systematically control the pressure for communicative efficiency, ensure that the data elicited from signers and gesturers are comparable, and facilitate comparisons in future research on other sign and spoken languages and across different population groups (e.g., children and adults). Thus, in this thesis I aim to go beyond the existing literature by not only assessing the role of modality-specific properties in achieving communicative efficiency but also proposing and implementing an experimental design that assesses communicative efficiency in a context of increasing information load during an interaction. I describe the design of the study in shortened form below; the full description can be found in **Chapters 2 and 3** for signers and **Chapter 4** for silent gesturers.

1.2.2. Methods

1.2.2.1. Participants

To answer the first two research questions, in **Chapters 2 and 3** I use data from the same participants—twenty-three deaf adults (12 female, M age = 30.5, range 18 – 57) who are native signers of LIS (N=17) or who acquired LIS in childhood (between ages 4-8 at school) in specialized schools for the deaf (N=6). Given the differences with regard to participants' age of acquisition, I control for this variable in the statistical analyses. To answer the third research question, in **Chapter 4** I use data from twenty-three hearing Italian adults (12 female, M age = 26.04, range 18–37) who are native speakers of Italian but who were asked to use only gestures to communicate (i.e., silent gesture). I then compare the data from the Italian speakers in silent gesture condition against the data from the LIS signers used in **Chapters 2 and 3**. An overview of the background and relevant characteristics of the deaf participants can be found in **Chapters 2 and 3**, and the hearing participants are described in **Chapter 4**.

Box 1-1. Historical background of LIS (Italian Sign Language)

During the Renaissance period, religious congregations in Italy ran numerous institutions for the education of deaf children and adolescents (Porcari Li Destri & Volterra, 1995). As such, implicitly, this setting fostered the spread and stabilization of the sign language used by the pupils (Porcari Li Destri & Volterra, 1995). The first institute for the deaf in Italy that integrated LIS in education was founded in Rome in 1784 by abbot Tommaso Silvestri (Geraci, 2015; Porcari Li Destri & Volterra, 1995). Since then, and up until 1880, numerous educational institutions for the deaf were established and the use of what was then considered gestural communication (and what is now known as LIS) was encouraged (Branchini, 2014).

The *International Congress of Education for the Deaf (ICED)* in Milan in 1880 marked the abrupt change in the education system of the deaf in Italy and other countries. Namely, a strictly oralist approach to education was proposed and adopted, while use of any sign systems was prohibited (Branchini, 2014). Nevertheless, deaf pupils in Italy would still use LIS outside the classrooms and within their communities. Furthermore, deaf adults began to form deaf associations, which eventually unified and became the National Deaf Organization (Ente Nazionale Sordi, ENS) in 1932 (Branchini, 2014; Volterra, 2011). ENS continues to be the main organization of the deaf community in Italy up to this day (Geraci, 2015).

The late 70's marked the beginning of the linguistic research on LIS carried out at the National Research Council in Rome. The first book that used the term *LIS* and was devoted to research on LIS was published in 1987 (Volterra, 1987). At the end of the 80's and beginning of the 90's, deaf groups offering LIS language and interpreting courses started to appear (Volterra, 2011). This was also the time when bilingual schools (from kindergarten to high school) offering education in LIS and Italian were established (Branchini, 2014). These schools still operate today (e.g., Istituto Statale di Sordi in Rome). While regular schools have been open to deaf children since 1977, a legal right to have a "communication assistant" who uses LIS (in case the child uses LIS) was granted only in 1994. Italy officially recognized LIS on May 19, 2021.

1.2.2.2. Design and material

For all three experimental studies presented in this thesis, I used an elicited task paradigm that I designed specifically for the present thesis in order to target and systematically assess simultaneous and iconic constructions in an interactive context which requires participants be communicatively efficient as information load increases. The design of the experiment consists of the systematic manipulation of the information density of an event, quantified as the specific number of semantic information units requiring encoding.

The stimuli presented were black-and-white cartoon images. Events represented in the images involved two animate referents and were manipulated in such a way that the information density of these events was systematically increased, requiring participants, whose task was to describe these images to another person (a confederate), to be more communicatively efficient as the information density of the events increased.

There were five information density levels in total (see Figure 1-6). The least informationally dense event consisted of two animate referents represented in a drawing as being next to each other, yielding 2 core information units requiring encoding, i.e., referent 1 and referent 2 (Fig. 1-6, level 1).

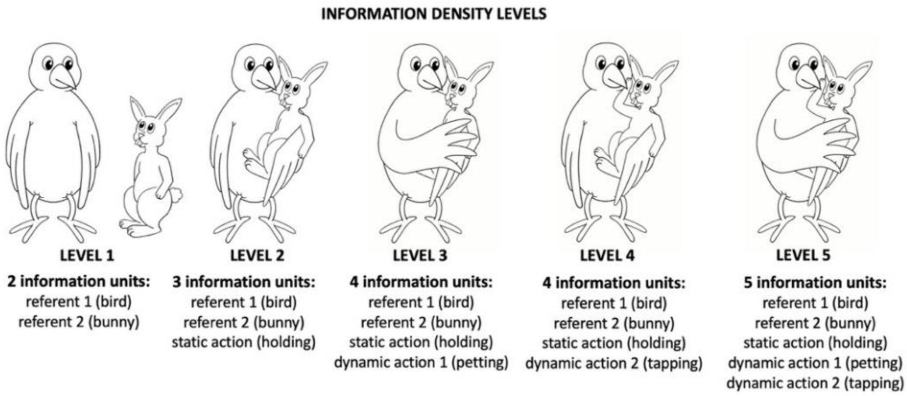


Figure 1-6. Example of 5 *Information density levels* for the referent pair: bird and bunny. The format of the images in levels 1 and 2 is PNG and the format of the images in levels 3, 4, and 5 is GIF. In the GIFs, the dynamic action of the referents is animated. For the original (animated) version of the stimuli, please see: https://osf.io/g57p2.

The most informationally dense event consisted of two animate referents interacting with each other, where referent 1 was holding and petting referent 2 while referent 2 was acting on referent 1. The action of referent 2 could be either petting, pinching, tapping, kissing, licking, or pecking, depending on the kind of animate referent represented. In the most informationally dense event, 5 core semantic information units required encoding – referent 1, referent 2, two actions by referent 1 and one action by referent 2 (Fig. 1-6, level 5).

To each consecutive information density level, one information unit was added. The exception to this were levels 3 and 4, which both had 4 information units. However, in level 3, referent 1 was agent of the action and referent 2 was patient of the action, while in level 4 both referents were agent and patient of the actions, thereby making the event represented in level 4 conceptually more complex than in level 3.

The materials for the stimuli were developed and designed in consultation with and with the assistance of the deaf research assistants and employees of the LaCAM laboratory at the Institute of Cognitive Sciences and Technologies – National Research Council Rome (ISTC-CNR Rome) and who were either native or highly proficient signers of LIS. One example of the fruits of this collaboration is the decision to use GIFs for the stimuli that involved dynamic actions (levels 3, 4, 5) in order to be able to animate the dynamic actions. This was done following joint discussion and pilot testing which revealed that interpreting the movement of the dynamic actions from still drawings can be problematic while animated movement provides sufficient clarity with regard to the specific action. All stimuli were designed and animated by the research assistant Barbara Pennacchi. There were six different animate referent pairs in total and each pair was represented at all five information density levels, yielding 30 images in total (for all images, see Appendix A). These images constituted the experimental stimuli used in all three studies.

The design and materials allowed me to assess the effect of information density of the event on simultaneous and iconic constructions as well as the linguistic strategies used in a systematic way given that the number of information units requiring encoding was fixed and controlled. Furthermore, it allowed me to use the same stimuli for both groups investigated in this thesis—signers and silent gesturers—

ensuring a direct comparison between these two groups. A complete overview of the experimental design and materials is presented in **Chapters 2, 3, and 4** because they are separately published papers.

1.2.2.3. Procedure

Each participant was told that they would play a *director-matcher* type of game and that they had been assigned the role of *director*. Deaf participants were instructed in LIS by a native LIS signer; hearing participants were instructed in Italian by a native or highly proficient Italian speaker.

The participant stood facing another person, who had the role of *matcher* (a confederate). When testing LIS signers, the matcher was a deaf person and when testing Italian speakers using silent gesture the matcher was a hearing person. The matcher sat behind a table, on top of which was a TV screen that faced the director (i.e., the participant) and a laptop that faced the matcher (i.e., confederate). The experimenter was seated at the left corner of the table and controlled from their laptop the presentation of the stimuli, which appeared on the TV screen. The participant was told that the director's task was to describe the images appearing on the TV screen to the matcher so the matcher could choose the correct image on the laptop.

As a warm-up, participants were presented with images of all the referents, one by one. Once the images were described to the matcher and the matcher had confirmed that they had made a selection on their laptop, the experimental stimuli were presented. They were displayed one by one in a semi-randomized order so that the same information density level and same referent pair never appeared one after another. Semi-randomization ensured that participants could not refer to the preceding image that might have matched with regard to the referents or information density level. Accordingly, participants were required to describe each stimulus independently. Once all images had been described, the participant was debriefed about the goal of the study and actual role of the matcher.

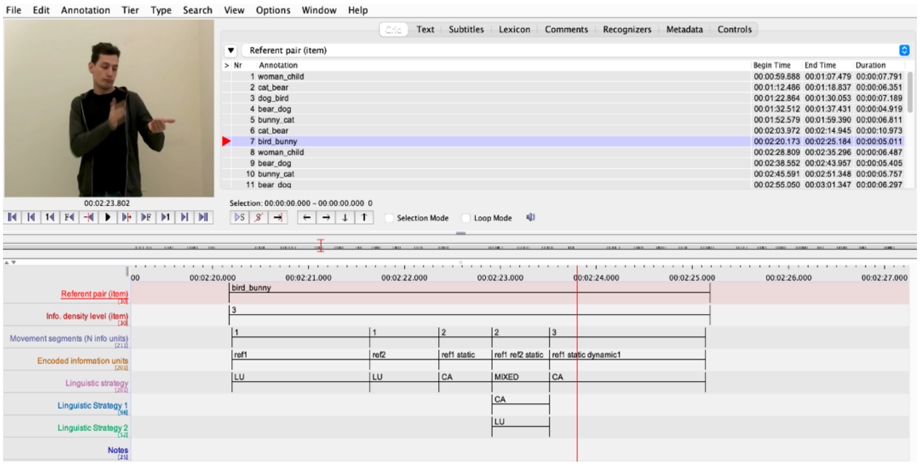
The participants' productions were video-recorded and used for coding. All participants signed consent forms before the start of the experiment, agreeing to be video-recorded and consenting that their data could be used for academic and scientific purposes. After the experiment was completed, participants received 5 EUR for having participated. A complete overview of the procedure is presented in **Chapter 2** for deaf participants and **Chapter 4** for hearing participants.

1.2.2.4. Coding

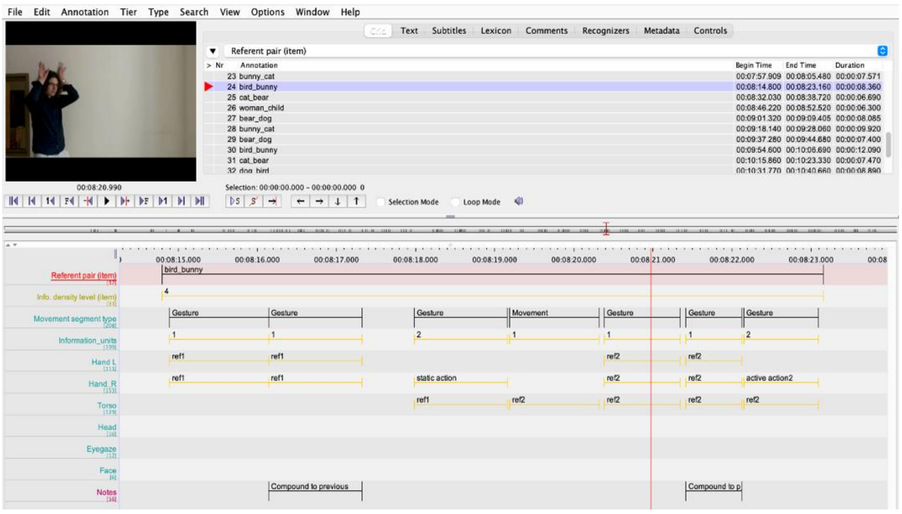
The video-recorded data were coded in ELAN, a free multimodal data annotation software developed by Max Planck Institute for Psycholinguistics (Wittenburg et al., 2006). Figure 1-7 presents screenshots of the coding in ELAN software for LIS data (a) used in **Chapters 2, 3, and 4** and silent gesture data (b) used in **Chapter 4**. All the productions from signers using LIS and from hearing participants using silent gestures were coded with regard to the information density level and referent pair of the experimental items.

LIS data were coded by the author of the thesis (hearing L2 signer of LIS). All coded data were later checked by a deaf researcher, a native signer of LIS. Another native signer of LIS independently coded 20% of the data for reliability measures. LIS data were coded for movement segments (MS), number of encoded information units in each movement segment and the information unit type (i.e., referent 1, referent 2, static action, dynamic action 1, dynamic action 2), as well as the linguistic strategy or strategies used in each movement segment (see Figure 1-7 a). The coded data regarding the length of the productions (number of movement segments used to describe an image) and simultaneous organization of information were analyzed in **Chapter 2.** The coded data regarding use of the specific linguistic strategies and their combination were analyzed in **Chapter 3**. For more details on the coding scheme and reliability for the LIS data, please see **Chapters 2 and 3**.

Silent gesture data were coded by two hearing native Italian speakers. The author of the thesis (L2 speaker of Italian) independently coded 20% of the data for reliability measures. Silent gesture data were coded for movement segments (MS), type of movement segments, number of encoded information units in each movement segment, and the type of the information unit encoded by each articulator (see Figure 1-7 b). Furthermore, data were coded for the use of compounded gestures (i.e., combination of 2 or more MS for encoding the same referent) during the encoding of information units involving the referents (ref. 1 and ref. 2). This additional step was relevant to ensure direct comparison with the sign language data. For more details on coding scheme and reliability regarding silent gesture data, as well as issues regarding the comparability of the silent gesture data with the LIS data, please see **Chapter 4**.



a



b

Figure 1-7. Screenshots of coding in ELAN software for LIS data (a) and silent gesture data (b).

1.2.3. Outline of the thesis and predictions

In the three empirical chapters of this thesis (**Chapters 2, 3 and 4**) that use the same experimental paradigm, I assess whether and how the visual modality-specific properties of simultaneity and iconicity are employed for efficient communication from different perspectives. In **Chapter 2** I assess whether the simultaneous organization of information is used for communicative efficiency (information organization perspective); in **Chapter 3** I assess what linguistic strategies are recruited for communicative efficiency and in particular the role of CA in achieving it (linguistic encoding perspective); finally, in **Chapter 4** I assess whether the use of simultaneous and iconic constructions constitutes an emergent linguistic property as opposed to simply a general affordance of the visual modality (language evolution perspective).

In **Chapter 2**, I assess whether signers of LIS use modality-specific properties of simultaneous and iconic constructions to encode events that are perceptually simultaneous in order to achieve communicative efficiency. I expect that if simultaneity and iconicity are used to cluster related meanings closer together to achieve communicative efficiency, signers of LIS will be likely to do so through the use of simultaneous and iconic constructions. Accordingly, I predict that when faced with increasing information demands signers will increase use of simultaneous and iconic constructions as well as the information density (i.e., number of semantic information units) encoded in a single simultaneous construction in order to cluster related meanings closer together. Alternatively, if the use of simultaneity remains constant or decreases as the amount of information requiring encoding increases, this may suggest that these properties are not recruited for communicative efficiency per se. I address these predictions by means of an experimental study in which I systematically manipulate the information density of the events that participants are asked to communicate to an interlocutor in an interactive task.

In **Chapter 3**, I use the same data as in **Chapter 2** to focus on the linguistic strategies, in particular on the role of the depicting strategy *constructed action (CA)* and how it is combined with other linguistic strategies (*depicting constructions, lexical signs, pointing*) to achieve communicative efficiency. I focus on CA because it allows the full spectrum of body articulators to be recruited in representing meaning through iconic mapping, and as a result CA allows for a direct and informative representation of an event. Furthermore, considering that this strategy employs both manual and non-manual articulators for the iconic mapping of meaning, this strategy can be combined with other linguistic strategies such as lexical signs and depicting constructions, which are articulated by the hands, allowing for more flexibility in expression. I hypothesize that if CA is used with referential intention in non-narrative context such as the one I used here and to achieve communicative efficiency, then its use either alone or in combination with other strategies should increase along with the increase in information demands. The aim of **Chapter 3** is to reveal whether CA, generally considered to be a discourse strategy used in narratives, can also be used outside of narrative contexts and with the function of achieving communicative efficiency.

In **Chapter 4**, I approach the communicative efficiency function of simultaneous and iconic constructions from an evolutionary perspective. Namely, I aim to assess whether simultaneous and iconic constructions used by signers constitute a general affordance of the visual modality or whether these constructions have evolved in a linguistic system as an adaptation for communicative efficiency. I attempt to address this consideration by comparing the use of simultaneity by the LIS signers analyzed in **Chapter 2** to the use of simultaneity by a group of hearing Italian adults with no knowledge of any sign language and who are asked to use only gestures to communicate (i.e., silent gesture). I hypothesize that if simultaneous and iconic constructions constitute an emergent linguistic property in sign languages that has been optimized for communicative efficiency, they should be taken advantage of more frequently and used in more sophisticated ways within a linguistic system (i.e., sign language) than outside of it (i.e., silent gesture). Alternatively, if silent gesturers use simultaneity to the same extent or even more than signers, this would indicate that an adaptation process for this property is not necessary, and as such it constitutes a general affordance rather than an emerging linguistic property. Thus, the aim of **Chapter 4** is to shift the focus from the emergence of linearization to simultaneity as a potentially emergent linguistic property of a linguistic system in the visual modality.

In **Chapter 5,** I bring together the results of the three experimental studies described in this thesis. I discuss the findings from all three studies in light of the theoretical implications for sign language processing, communicative efficiency, and language evolution, as well as highlight avenues for future research.

Chapter 2

The role of iconicity and simultaneity in efficient communication: The case of LIS (Italian Sign Language)

This chapter is based on:

Slonimska, A., Özyürek, A., & Capirci, O. (2020). The role of iconicity and simultaneity for efficient communication: The case of Italian Sign Language (LIS). *Cognition, 200*, 104246.

Abstract

A fundamental assumption about language is that, regardless of language modality, it faces the *linearization problem,* i.e., an event that occurs simultaneously in the world has to be split in language to be organized on a temporal scale. However, the visual modality of sign languages allows its users not only to express meaning in a linear manner but also to use iconicity and multiple articulators together to encode information simultaneously. Accordingly, in cases when it is necessary to encode informatively rich events, signers can take advantage of simultaneous encoding in order to represent information about different referents and their actions simultaneously. This in turn would lead to more iconic and direct representation. Up to now, there has been no experimental study focusing on simultaneous encoding of information in sign languages and its possible advantage for efficient communication. In the present study we assessed how many information units can be encoded simultaneously in LIS (Italian Sign Language) and whether the amount of simultaneously encoded information varies based on the amount of information that is required to be expressed. Twenty-three deaf adults participated in a *director-matcher* game in which they described 30 images of events that varied in the amount of information that they contained. Results revealed that as the information that had to be encoded increased, signers also increased their use of multiple articulators to encode different information (i.e., kinematic simultaneity) and density of simultaneously encoded information in their production. Present findings show how the fundamental properties of sign languages, i.e., iconicity and simultaneity, are used for the purpose of efficient communication in LIS.

2.1. Introduction

In order to share a thought with others through language we need to decompose our message into smaller information units and then organize these units on a linear scale. While all languages face this *linearization problem* (Levelt, 1980, 1981), the impact of it might be different based on the main modality the language is realized in. "For spoken language, linearization is an absolute requirement" (Levelt, 1980, p.153) given that a thought has to be split in a strictly sequential manner (i.e., sounds arranged in words, words - in sentences, sentences - in discourse etc.).[4] Although Levelt (1980, p.156) speculates that the linearization problem might also stand for sign languages as it does for spoken languages, empirical research on this subject and to what extent not only linearity but also simultaneity is fundamental for communication for sign languages is scarce. In sign languages, affordances of the visual modality allow packaging multiple units of information not only sequentially, but also simultaneously, through the use of multiple articulators (hands, torso, head, eye gaze and facial expression). Simultaneity of the articulators also allows messages to be expressed in an iconic manner (i.e., *diagrammatic iconicity*, Perniss, 2007a; Risler, 2007; Taub, 2001).

Following the cognitive linguistics approach, language is tightly interrelated with general cognitive processes of the human mind (Croft et al., 2004; Elman et al., 1996). The linguistic signal is fast and fleeting in both the acoustic and visual modalities. This implies that information has to be encoded efficiently not only regarding the effort of the producer but also in relation to the informativity for the perceiver to understand it (Gibson et al., 2019; Grice, 1975). Thus, encoding information faces the communicative efficiency problem in both the acoustic and visual modalities. To what extent using simultaneity of the articulators is recruited to express information simultaneously and iconically for efficient communication in sign languages is underexplored.

[4] Note that some simultaneity in spoken communication is possible when taking into account that information can be conveyed in gesture and/or prosody alongside speech (Clark, 2016; Kendon, 2001, 2014; McNeill, 2008; Wilson & Wharton, 2006).

In this paper, we are interested in how the capacity of sign languages to encode multiple semantic information units simultaneously is modulated when producing informative messages of different information density in LIS (Italian Sign Language). We hypothesize that in sign languages efficient communication about perceptually simultaneous events can be achieved by strategic use of simultaneous and iconic encoding of information. We predict that signers will make more use of simultaneous and iconic constructions as a function of the amount of information that is required to be encoded for efficient communication.

In the following paragraphs, we will describe first how efficient communication has been addressed in spoken languages and then proceed to elaborate the notions of iconicity and simultaneity in sign languages. We will then test whether signers take advantage of simultaneous and iconic constructions when encoding informationally dense messages in a semi-naturalistic experiment.

2.1.1. Information packaging for efficient communication

The pressure to produce and process information rapidly has far-reaching implications for how languages are organized in regard to their structure (Christiansen & Chater, 2016; Hawkins, 2004; Jaeger & Tily, 2011; Lu et al., 2016). In spoken languages efficient communication at a syntactic level is achieved by clustering related meanings closer together, i.e*., dependency distance minimization* (Gibson et al., 2019; Hawkins, 2004) and reducing linguistic forms when possible (Aylett & Turk, 2004; Jaeger, 2010; Jaeger, 2006; Jaeger & Tily, 2011; Levy & Jaeger, 2007; Mahowald et al., 2013; Piantadosi et al., 2011). Considering that language encoding and decoding is dependent on cognitive constraints common to all language users regardless of the linguistic modality (Christiansen & Chater, 2016), it is plausible to assume that a similar strategy like minimizing dependency distances would be also used in sign languages for achieving efficient communication. On the other hand, efficiency in regard to reduction of linguistic forms in spoken languages has been mainly concerned with reduction and omission of function words (Ferreira & Dell, 2000; Jaeger & Tily, 2011; Levy & Jaeger, 2007; Race & MacDonald, 2003; Tagliamonte et al., 2005) which in sign languages mostly do not exist (Fang et al., 2007). Thus, this strategy is less likely for sign languages. For this reason, in this paper we focus on dependency distance minimization only.

It has now been established that users of spoken languages tend to cluster words that are syntactically and semantically related (Bybee, 1985, 2013; Futrell et al., 2015; Gibson, 1998, 2000; Gildea & Temperley, 2010; Grodner & Gibson, 2005; Hawkins, 2004; Liu et al., 2017; Temperley, 2007; Temperley & Gildea, 2018). It is argued that this strategy reduces cognitive load on working memory when forming relations among adjacent words in both production and comprehension (Ferrer i Cancho, 2004; Ferrer i Cancho & Liu, 2014; Gibson et al., 2019; Hawkins, 2004; Jiang & Liu, 2015; Liu, 2008). By means of dependency distance minimization, efficient communication can be achieved given that access to syntactic and semantic representation is provided as fast as possible (Hawkins, 2004, p.9) both for producers and comprehenders.

It would be expected that encoding syntactically and semantically related information closer together also in sign languages would lead to reduction of cognitive load as in spoken languages. Given that sign languages allow encoding related information not only sequentially but also simultaneously, it is plausible to assume that simultaneous encoding of information could be exploited for achieving efficient communication. Accordingly, the differences in affordances of language modality could potentially lead to different strategies of how information is packaged for efficient communication.

In the next section we describe the role of iconicity and simultaneity for direct (i.e., with motivated form-meaning mappings) representation of complex events and argue that use of simultaneous encoding of information could be potentially used for achieving efficiency in communication.

2.1.2. Iconicity and simultaneity in sign languages

The role of iconicity i.e., a structure-preserving mapping between mental models of linguistic form and meaning" (Taub, 2001, p. 23), has become a hot topic in language research in recent years and has been finally brought into prominence as a general property of language (Dingemanse et al., 2015; Perniss et al., 2010; Perniss & Vigliocco, 2014). Iconicity plays a fundamental role in language from the very beginning of its development as it provides motivated links to the experience in the world (Ortega, 2017; Ortega et al., 2017; Perniss & Vigliocco, 2014). Iconicity can be expressed at the lexical level, i.e., encoding correspondence between visually perceived features of the sign and its referent (e.g., sign for *a bird* resembles a beak of the bird in LIS), called *imagistic* iconicity (Russo, 2004; Taub, 2001). Similarly, also in spoken languages we can find words that perceptually resemble their meaning like onomatopoeias *meow* and *bang* in English or ideophones *sinisinisini* "closely woven", *saaa* "cool sensation" in Siwu (Dingemanse, 2011, 2015).

Another type of iconicity found in both sign and spoken languages is called *diagrammatic iconicity* (Haiman, 1985; Perniss, 2007; Pietrandrea & Russo, 2007; Taub, 2001; Ungerer, 1999). This type of iconicity is not linked to perceptual but to structural resemblance between form and meaning. That is, the relationship between parts of a specific meaning motivates the relationship between parts of its linguistic form. For example, structural resemblance can be identified in the previously mentioned sign *bird.* In LIS it is performed by a closed fist and extended thumb and index finger that open and close. This sign is positioned in front of the signer's mouth with fingers facing outwards. Here we see that even the smallest building blocks of the sign, i.e., parameters: handshape, location, orientation & movement are not meaningless units but instead have imagistic iconic properties (Boyes-Braem, 1980; Emmorey, 2014; Occhino, 2017) while their relation to each other is diagrammatically iconic (Emmorey, 2014; Lepic & Occhino, 2018; Taub, 2001). Furthermore, diagrammatic iconicity is also present in compounding when two signs are combined or merged for a new meaning, e.g., in ASL sign *inform* consists of two signs *know* and *offer* articulated in a single smooth movement (see Lepic, 2015 for this and other forms of compounding in ASL). In spoken languages, too, we can observe diagrammatic iconicity in compounding (Lepic, 2015; Ungerer, 1999), e.g., in English, the word *glamping* consists of two words, *glam(orous)* and *camping,* which have to be interpreted in relation to each other in order to derive the meaning of the compounded word (Lepic, 2015, p.212).

Diagrammatic iconicity goes beyond the lexical level and allows for the establishment of meaningful semantic and syntactic relations. In spoken languages diagrammatic iconicity has been acknowledged as an integral part of grammatical structure (Haiman, 1985). However, it is mainly limited to the iconicity of sequence as in "the order of statements in a narrative description corresponds to the order of the events they describe" (Haiman, 1980, p.516). In spoken languages diagrammatic iconicity is strictly linear considering that "since it is impossible to say everything at once, words must appear in a certain order" (Haiman, 1980, p.528). When it comes to sign languages, the strictly linear nature of information unfolding can be overcome. Given their visual modality and accordingly the use of not only linear but also three-dimensional space for linguistic encoding, diagrammatic iconicity can be taken advantage of in order to sign multiple meaning elements simultaneously[5], while establishing motivated relations between them (Risler, 2007). For example, events that are perceived simultaneously in the world, e.g., events involving multiple referents and/or their actions and interaction, can be encoded in such way that perceptual simultaneity consisting of multiple meaning elements (i.e., information units) is encoded simultaneously in linguistic form as well. This can manifest in different ways depending on the type and complexity of the event.

For example, in order to encode an agent, their action/s and a patient, e.g., a woman holding a dog, a signer after introducing both referents with lexical signs (*woman, dog*) could then take on the role of the woman (agent) by mapping the woman onto the signer's body (e.g., through facial expression, eye gaze, and/or torso) and at the same time encode the holding action by one of the hands representing holding (Figure 2-1a). Note that in this example two schematized elements of the same event are encoded, i.e., the agent and the action in respect to the agent. The dog is only available implicitly while the woman and the holding action are encoded explicitly. If the event to be communicated consisted of a woman holding AND petting the dog, the signer could simply superimpose another information unit (i.e., *petting*) by doing the petting action with the other hand in respect to the location of the referent it should

[5] Note that diagrammatic iconicity can be also used sequentially for encoding topographic and static relations between referents (Emmorey, 2002; Emmorey & Tversky, 2002; Özyürek et al., 2010; Perniss et al., 2015; Sümer et al., 2013). Referents can be encoded and positioned in space consecutively, but they are nevertheless "conceived as present" in conceptual representation, e.g., signing different types of furniture and positioning them in space one after another to encode the layout of a room.

be acting on, in this example - a dog being held by the woman (Figure 2-1b). A signer mapping the referent onto their body and producing actions in relation to this embodied representation is known as *constructed action (CA)* (Cormier, Smith, et al., 2015; Metzger, 1995; Tannen, 1989).

a b

Figure 2-1. A signer encoding a) a woman who is holding a dog where the woman is mapped onto the body of the signer marked by torso, head, eye gaze, and facial expression, and the holding action is encoded on the right hand, and b) superimposing an additional action of petting (left hand) onto the previously encoded representation of the woman holding the dog.

Although not acknowledged in previous literature such constructions involve not only imagistic properties of individual lexical items but diagrammatic iconicity as well. Namely, in such constructions each articulator can be used to encode different information about a referent simultaneously while being in a meaningful relation to each other through iconic schematization as exemplified above.

Alternatively, it would also be possible to use constructed action to encode information units consecutively, e.g., first encoding the holding action only and then encoding the petting action only in a location where the signer had just encoded the holding action. In this scenario, diagrammatic iconicity would be maintained, considering that both actions are linked with motivated use of space (they are performed based on the location and properties of the patient). However, simultaneity would be lost and accordingly a single representation would be split in two parts (first holding, then petting). These two parts would nevertheless be iconically linked to each other. Finally, encoding could be also achieved using specific lexical signs (i.e., *to hold*, *to pet*) articulated in neutral space in front of the signer and thus without resorting to diagrammatic iconicity at all. In this scenario, each information unit would be presented separately from the others without establishing an iconic link between the signs. As a result, such encoding would form neither single nor split iconic representation from explicitly encoded information as in the previous two cases. Such encoding would rely on maintaining previously presented information in working memory only and it would be interpreted exclusively based on specific linguistic structure as in spoken languages.

A more complex event could involve two referents interacting with each other. In this scenario, mapping one of the referents and its actions onto the signer's body would encode only part of the event. The action/s of the other referent would then need to be encoded in a consecutive manner. It is then possible for signers to use a set of articulators to encode one referent and/or their actions, while other articulators can be used to encode another referent/s and/or actions. Such encoding can be achieved through *body partitioning* (Dudis, 2004). Dudis (2004) discusses partitionable zones of signer's body that allow the production of distinct information elements (i.e., information units) directly accessible to interlocutors. As main partitionable zones Dudis considers both hands and face. The face can be further divided into 2 zones that can be used to express different information units – facial expression (including eye gaze) and oral articulators (e.g., mouth). Given that signers can partition their body to create distinct information units it is possible to encode information involving multiple referents simultaneously. For example, Pizzuto et al. (2006) describe a signer encoding the boy holding a dog and the dog licking the cheek of the boy in LIS (Figure 2-2). In this example, there are 4 main semantic information units (*distinct elements* in terms of Dudis, 2004) that are explicitly encoded, namely 1) *the boy* who 2) *is holding* 3) *a dog* and the dog who 4) *is licking* the boy. The face of the signer is partitioned in such way that the upper part of the face including eye gaze is the boy gazing at the dog while the facial expression including the mouth refers to the dog (thus, we have evidence that eyes can be partitioned from the rest of the face as well). Note that the head is to be considered that of the boy because the sign for *licking* is produced on the cheek of the signer, i.e., the cheek of the boy who is looking at the dog. The lexical sign for *licking* (right hand sign) can be diagrammatically positioned in respect to the constructed action of the boy holding the dog. Here we can observe how iconicity affords integration of multiple elements that can be combined to create a single representation of the event. If we take into account the eye gaze and location of licking as crucial information that has to be encoded then we have 6 information units (the boy, holding, the dog, licking, the cheek, and eye gaze direction of the boy) available at the same time.

Figure 2-2. A signer encoding a boy holding a dog and the dog licking the boy. The boy is mapped onto the body of the signer and marked by torso and eye gaze. Boy's holding action is encoded on the left hand. The dog is marked by facial expression of the signer and mapping between the dog's mouth and the signer's mouth. Licking action is encoded on the right hand and mouth of the signer. Figure has been recreated for demonstration purposes (see original in Pizzuto et al., 2006, p.483).

To summarize, in sign languages different articulators encoding different information units are all interpreted in relation to each other in given moments in time. The interpretation of such simultaneous constructions is possible due to diagrammatic iconicity that establishes these relations and links them to semantic representation (Risler, 2007, p. 75). In the next section we briefly review previous research on how the amount of simultaneously encoded information has been previously studied.

2.1.2.1. Simultaneous information encoding in sign languages

In research on simultaneity we can encounter different kinds of terminology that refers to different phenomena. For this paper, it is important to differentiate between *encoding of simultaneity* and *simultaneous encoding*. While the former refers to the grammatical structure of encoding referents and/or actions that appear/happen at the same time by using various linguistic strategies (in both spoken and sign languages) the latter refers specifically to the property of sign languages to encode different semantically or discourse-related information simultaneously by means of simultaneous constructions. The present paper focuses on the latter.

Due to the affordances of the modality spoken languages encode simultaneity in a linear manner by using specific lexical items (e.g., while, and, as) to link pieces of meaning units together to transmit information that is semantically simultaneous (Morgan, 2002). The same applies if speakers want to elaborate further on the same event. In spoken languages, this has to be done sequentially. Also, in sign languages it is possible to use lexical markers or sequential grammatical structures like "sandwiching", which is the doubling of the verb, to indicate simultaneity of the events (Fischer & Janis, 1990; Morgan et al., 2002; Napoli & Sutton-Spence, 2014). Alternatively, due to use of multiple articulators and signing space simultaneous encoding can be employed (Morgan et al., 2002; Napoli & Sutton-Spence, 2010). In the present paper we are concerned with simultaneous encoding of information related to the events that are perceived simultaneously.

Note that simultaneous encoding in sign languages is possible not only for events involving multiple referents engaged in simultaneous activities but also for the information about the same referent like encoding different perspectives of the same referent/action (Perniss, 2007a, 2007b). Simultaneous encoding can be also used to mark other properties of the discourse (e.g., various types of listing, Liddell et al., 2007). Vermeerbergen et al. (2007) devote an entire volume to simultaneity, the first and currently only volume devoted to this topic. They differentiate between specific combinations of articulators used for specific purposes: manual simultaneity where each hand encodes different information; manual-oral simultaneity where mouthing's and mouth gestures contribute to the meaning expressed by the hands; simultaneous use of manual and non-manual articulators to encode different perspectives of the same event or multiple events simultaneously. In the present paper, we are not concerned with the grammatical structure of encoding of events that are perceived simultaneously, but rather with the amount of information that can be packaged simultaneously to encode such events.

There are some descriptive studies that assess the capacity and/or limits of the amount of information that can be encoded simultaneously in a sign language (Dudis, 2004; Napoli & Sutton-Spence, 2010; Perniss, 2007a, 2007b; Sutton-Spence & Boyes Braem, 2013). Note that all research on simultaneous information encoding has been exclusively based on narrative data and has been mostly descriptive with little to no consideration of the functional role of iconicity in linking different information units into single representation. One hypothesis, however, has been proposed in the past. Namely, that simultaneous and iconic constructions are used for communicative efficiency and informativeness. Perniss (2007a, 2007b) explores simultaneous encoding of the same event from different perspectives, observer and character perspective (i.e., *non-prototypical alignment*), and proposes that such simultaneity is used to achieve communicative efficiency and informativeness for coherence and ambiguity reduction in discourse. However, this research is also based on narrative data. This factor may create an issue in assessing the communicative function of simultaneous information encoding in signing. Namely, narrating a story requires a signer to be not only informative (i.e., referential function) but also expressive and entertaining (i.e., evaluative function) in their production (Labov & Waletsky, 1967; Özyürek & Trabasso, 1997). It has been argued that a highly iconic strategy like constructed action, which is directly linked to simultaneous information encoding as described in previous paragraphs and to *non-prototypical alignment*, is used to fulfill the evaluative function in narratives as opposed to the referential function (see Rogers, 2012 for a review). Accordingly, it is impossible to tease apart whether simultaneity is used as a narrative device to enhance the evaluative function of narration or whether it is used to increase communicative efficiency.

In relation to communicative efficiency, description of simultaneity is mostly concerned with encoding the same information from different perspectives as described above (Perniss, 2007a, 2007b) or alternatively with simultaneity of multiple predicates (Napoli & Sutton-Spence, 2010; Risler, 2007). Accordingly, there is no consideration that a single predicate might contain multiple semantically distinct information units which are encoded simultaneously and linked through diagrammatic iconicity.

Napoli & Sutton-Spence (2010) assess limitations on simultaneity in ASL (American Sign Language) by counting how many "propositions"[6] can be encoded simultaneously. Authors note that simultaneous encoding of two propositions is quite common in everyday signing while simultaneous production of three or four propositions is found in creative signing like poems. Authors find that a maximum of four propositions can be encoded simultaneously. Risler (2007), similarly like Napoli & Sutton-Spence (2010), accounts for predicative simultaneity and describes how two processes[7] can be encoded simultaneously in LSF (French Sign Language). As a result, the full range of simultaneously encoded information is not accounted for. For instance, in the example described in Figure 2-2, Napoli & Sutton-Spence (2010) would only account for two simultaneous "propositions", and Risler (2007) would account for two "processes", disregarding the richness of the information encoded if each semantic information unit is considered in its own right. To our knowledge, simultaneously encoded distinct information, regardless of its grammatical status, available to the interlocutor has been described only by Dudis (2004)[8] as elaborated in the previous section (see Section 2.1.2.). Surprisingly, out of all research on the amount of simultaneously encoded information, only Risler (2007) explicitly acknowledges the central role of iconicity by accentuating the importance of the diagrammatic link between arrangement of articulators and their semantic representation.

---

[6] Napoli & Sutton-Spence (2010) define *proposition* as follows: "[..] a proposition is a predicate and its constellation of arguments (Johnston et al., 2007) and is free of internal conjuction." (p. 650)

[7] Process signs are defined as "signs that express processes (actions or events)" (Risler, 2007, p .73)

[8] In his article, Dudis (2004) follows the theory of "conceptual blending" by Fauconnier & Turner (1996) and accordingly discusses the amount of visible distinct elements projected into real-space blends in signed production (p. 225). In the present study, we rephrase the phenomenon described by Dudis as "simultaneously encoded information".

To summarize, diagrammatic iconicity constitutes a tool in sign languages for packaging semantically related information which is particularly useful when dealing with informationally dense events involving multiple referents and their interactions. Instead of encoding each information unit sequentially, signers can construct a single representation or superimpose new information on already encoded information. As a result, multiple related information units are explicitly accessible in encoding and decoding. It is then logical to assume that more direct mapping of the event through simultaneity would lead to more efficient encoding in comparison to strictly sequential encoding. However, while previous research provides understanding of how simultaneous information encoding can be achieved and suggests that it might be used for efficiency and informativeness, there is no direct experimental evidence that this unique property of sign languages to encode related information simultaneously is actually used for such purpose.

2.2. The present study

One would expect that if dependency distance minimization lies at the heart of efficient communication, language should adapt to promoting related information being found as close together as possible (Hawkins, 2004). Up to now, this has been studied in spoken languages (i.e., dependency distance minimization) but not in sign languages. Exploring in a systematic way how simultaneity is used in sign languages could shed light on whether this information organization strategy is used for efficient communication. While there is some descriptive research available on how simultaneous encoding of information is achieved and how much information can be encoded simultaneously as discussed above, it is not clear when signers would use simultaneous constructions.

In the present study, we aim to test whether signers use simultaneous and iconic information encoding as a strategy to achieve efficient communication, specifically in LIS. We hypothesize that the more simultaneously perceived information related to the same event has to be encoded, the more simultaneous constructions will be used to cluster related information closer together. Furthermore, we hypothesize that the more information that has to be encoded, the informationally denser simultaneous constructions will be (i.e., more information units encoded simultaneously) in order to achieve a more direct representation of the event. We test our hypotheses by means of an elicited task paradigm in which we assess how the amount of information that has to be encoded, influences the amount and information density of simultaneity of encoding in deaf signers of LIS. Considering that previous research on simultaneity has mainly described simultaneous encoding of events involving actions of animate referents, we assess how signers use simultaneous information encoding when dealing with the same kind of events. However, unlike extracting examples from narrative data as has been done before, in which simultaneous constructions can be driven by many different factors and which are not directly comparable to each other, we construct single image stimuli that vary systematically in regard to how many semantic information units they contain. We then present these stimuli to the participants in a context of an interactive game in which they have to describe the images in order for the other person to choose the correct image. Accordingly, in the present task signers are faced with the necessity to be as informative and as clear as possible but without additionally enhancing the description with evaluative properties that are typical for narratives. As a result, we can directly test how signers manipulate the use of simultaneous constructions when faced with a task of encoding perceptually simultaneous events that systematically vary in their information density.

2.3. Methods

The study has been approved by the Ethics Council of the National Research Council of Italy (protocol n. 0012633/2019).

2.3.1. Participants

Twenty-three deaf adults (12 female, M age = 30.5, range 18 - 57) participated in the study. Seventeen participants were native signers of LIS, all children of deaf parents. Six participants were children of hearing parents and acquired LIS between ages 4-8 at school. All participants reported using LIS daily as their main language. We account for differences in age of acquisition of LIS in the analyses. Participants were recruited via a mailing list available to The Institute of Cognitive Sciences and Technologies and via a recruitment video created in LIS posted on various social media sites. All participants signed consent forms agreeing to be video-recorded and consenting that their data could be used for academic and scientific purposes. For their participation, participants received 5 EUR.

2.3.2. Design

In our design, we systematically increased the information density[9] of the events that have to be encoded. In the first level only two referents (i.e., two information units) had to be encoded. Participants could encode them by positioning referents in space or also by simply naming them. In the second level, participants had to encode both referents and one static action; in the third level - two referents, one static action and one dynamic action of the same referent (dynamic action 1); in the fourth level – two referents, one static action of one referent and one dynamic action of the other referent (dynamic action 2). Finally, in the fifth level the participants had to encode two referents, one static action and two dynamic actions of both referents. Below we describe the information density levels in more detail:

- ***Information density level 1*** (two referents = two information units in total) (Fig.2-3 - level 1):
  The least dense scenario in our design required referring to two animate referents. For example, a bird and a bunny. This information constituted the first density level in the design where two information units were required to be encoded.

- ***Information density level 2*** (two referents + 1 static action = three information units in total) (Fig.2-3 - level 2):

[9]*Information density* in this study is quantified as the total number of information units per experimental item (i.e., an image depicting an event). The term *information* in this study is used to refer to distinct semantic meaning units. Accordingly, we do not draw any parallels with information density of the Uniform Information Density framework (Jaeger, 2006; Levy & Jaeger, 2007) where information is defined in an "information-theoretic sense—the negative log-probability of an event [..]" (Levy & Jaeger, 2007 p. 849) and information density is defined as "the amount of information per unit comprising the utterance" (p. 849).

Then, we increased the information density by one unit by attributing a static action of holding to one referent in relation to the other referent that is being held. For example, the bird holding the bunny. Accordingly, in level 2, three information units were required to be encoded.

- ***Information density level 3*** (two referents + 1 static action + 1 dynamic action = four information units in total) (Fig.2-3 – level 3):

  Next, we increased the information density by one more unit by adding a dynamic action to the referent doing the static action. For example, the bird holding and petting the bunny. This resulted in 4 information units that had to be encoded.

- ***Information density level** 4* (two referents + 1 static action + 1 dynamic action of other referent = four information units in total) (Fig.2-3 - level 4)**:**

  Then, we shifted the agent of the dynamic action, e.g., instead of the bird holding and petting the bunny, the bird was holding the bunny and the bunny was tapping the cheek of the bird. In this level the bird became not only the agent but also the patient of the action. Note that here the information density did not vary from the previous level (we took away one action from one referent and added one to the other). However, here referents were both agents and patients simultaneously, thus creating a perceptually more complex event and a need to refer to both referents in order to encode both actions in contrast to the previous level where only one referent of the action had to be identified. Accordingly, while we did not manipulate information density between levels 3 and 4, we manipulated the complexity between these two levels.

- ***Information density level 5*** (two referents + 1 static action + 2 dynamic actions = five information units in total) (Fig.2-3 – level 5):

  Finally, we increased the information density of the event even further by including dynamic actions by both of the referents, e.g., the bird holding and petting the bunny while the bunny is tapping the cheek of the bird. In this level, 5 information units had to be encoded.

In encoding of all levels, signers could use different strategies of information encoding - sequential, simultaneous and mixed. Note, that it is impossible to encode the message without first introducing the referents via lexical signs, thus resulting in presence of linearity on all levels. However, the encoding of the event itself is more flexible and could potentially lead to both sequential and simultaneous encoding.

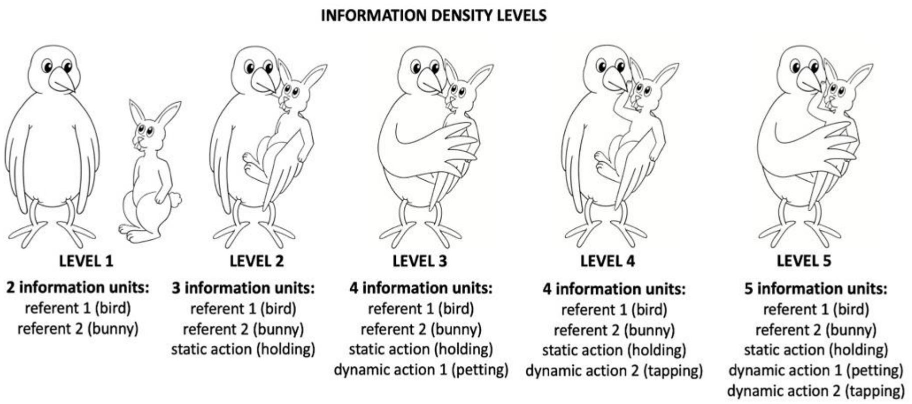


Figure 2-3. PNG stimuli of information density level 1 and level 2 and GIF stimuli of information density level 3, level 4, and level 5 (referent pair: bird and bunny). In GIFs dynamic action of the referent 1 and dynamic action of the referent 2 are animated.

2.3.3. Material

The material for the experiment consisted of 6 sets of 5 stimuli (PNG images for levels 1 and 2 and GIFs for levels 3, 4, and 5) representing each information density level in each set (see Fig. 2-3). We decided to use GIFs for the stimuli that involved dynamic actions based on the pilot studies which revealed that interpreting the movement of the dynamic actions from still drawings was problematic.

Thus, a total of 30 unique combinations of referents and actions were prepared for the experiment (see Appendix A). In each experimental trial, two animate referents were represented. In order to reduce bias for signers to personify with animate referent that is a human, both referents were animals with exception of one pair in which both referents were humans (a woman and a boy). In picture sets, animals alternated between referent 1 and referent 2. Referent 1 was always the bigger referent represented on the left side of the image. Referent 2 was the smaller referent represented on the right side of the image. Note that in all the stimuli the eye gaze of the referents was kept constant - both referents were looking at each other throughout. All experimental stimuli are freely available online (https://osf.io/g57p2/).

2.3.4. Procedure

A participant was greeted by an experimenter (a deaf researcher) and informed that the participant is about to play a *director-matcher* game with another deaf person. The experimenter also noted that all instructions would be given via video-recording once the experiment started. The participant was standing in front of another player who was seated in a chair in front of a table with a laptop (see Figure 2-4). This person was a confederate and not a naïve player. It was necessary to use a confederate for the task in order to ensure that all participants would receive comparable feedback. Previous research has shown that participants tend to adjust their communicative strategy based on the interlocutor's feedback (Holler & Wilkin, 2011). The confederate was instructed to always provide positive feedback (i.e., a head nod, *OK* sign, *yes* sign, *got it* sign) and no signals of doubt once the participant has finished their description of the stimuli.

On the left side of the table there was a 40-inch screen in which all instructions and stimuli were presented and which was not visible to the matcher. The experimenter was seated at the left side of the table and controlled the presentation of the video and stimuli by means of the laptop connected to the TV screen.

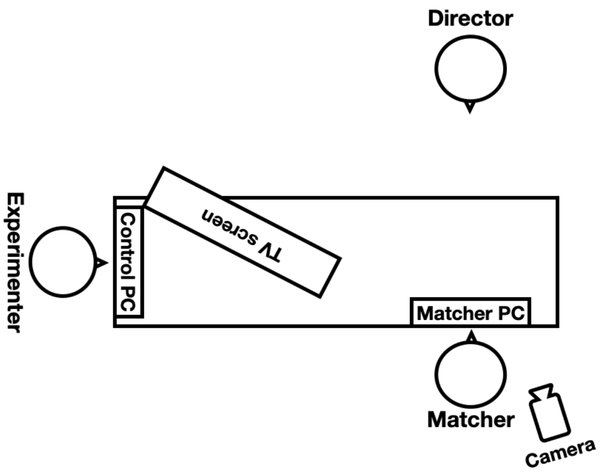


Figure 2-4. Experimental setup. Camera located on the tripod (approx. height - 1,50m)

When the participant was standing in the right place the experimenter started the instruction video. The participant received an instruction (via video-recording in LIS) that they were about to play a *director-matcher* game and that they were assigned the role of *the director*. They were informed that the matcher had multiple images presented on their laptop. The task of the director was to describe images so that the matcher could choose the right one. Participants were asked to look carefully at the screen, memorize the image/GIF and then sign it to the matcher. A pilot study revealed that sometimes it was difficult for the participants to remember the images, thus we decided to leave the images on the screen in case there was a need for the participants to double-check. However, they were explicitly asked not to look back at the screen while signing but to face the matcher in order to make comprehension easier for the other player.

First, participants were informed that before the start of the actual game they would see all the referents that have been picked for the game and that they should identify these referents to the matcher to make the guessing of the referents in the actual game easier. Thus, participants could describe each referent separately where they could also be able to provide details about their looks and physical aspects and introduce them to the addressee. We chose to present all referents separately first with a consideration that this task would make it less likely that participants would concentrate on physical details of the referents during the experiment and will instead concentrate on the focus of the study – encoding events depicted in the images. As a result, we would have cleaner data. Each referent was presented via PowerPoint presentation, one by one. When the director had named a referent and the matcher had nodded to indicate that they have understood and had picked an image on their laptop the experimenter proceeded to the next image. This procedure also provided a warm-up session and a grasp of the game.

Once all referents have been named, the experimenter announced that all the referents of the game have now been revealed and the game itself can start. If the participant had no further clarification questions, the experiment started. The process was the same as with the naming of images with single referents. If the participant had omitted an action (e.g., in *Information density level 4*, the participant forgot to mention that referent 1 is holding referent 2 but only encoded the dynamic action of referent 2), the experimenter asked them to watch the stimuli again carefully and to repeat their production. All stimuli were presented in a semi-randomized order. We randomized stimuli in such a way that the same referent pair did not appear one after another, also the same information density level did not appear one after another. Accordingly, it was necessary to encode all information units depicted in the images rather than contrasting only specific features. No contrasting strategy appeared in our data and each image was always described independently from other images. The productions of the participants were video-recorded and used for coding.

2.3.5. Coding

The video-recorded data was coded in the multimodal data annotation software ELAN developed by Max Planck Institute for Psycholinguistics (Wittenburg et al., 2006). The duration of the videos was 7.23 min on average (SD = 1.39).

We developed a coding scheme that enabled us to assess the simultaneous productions of the participants. For each stimulus (annotated with its information density level and referent pair) we coded:

Length of the encoding – the total number of "movement segments" per production.

Kinematic simultaneity - Simultaneous versus non-simultaneous use of manual and/or non-manual articulators in each movement segment.

Information density of simultaneity – the total number of encoded information units in each movement segment:

1 information unit (e.g., referent 1)

2 information units (e.g., referent 1 + referent 2)

3 information units (e.g., referent 1 + referent 2 + dynamic action of ref.1)

4 information units (e.g., referent 1 + referent 2 + dynamic action of ref.1 + static action of ref.1)

5 information units (e.g., referent 1 + referent 2 + dynamic action of ref.1 + static action of ref.1 + dynamic action of ref.2)

2.3.5.1. Movement segments and length of encoding

To determine the sequential organization of each production we segmented data into movement segments (MS) based on the start and end of a movement of the hand/s: i.e., segmentation of *a stroke* as defined by Kendon (2004). However, in our coding (differing from Kendon) a movement segment could include not only the stroke produced by the hand but also the hold of the previous sign if it remained present during the new stroke (see Figure 2-5, MS3-5). Thus, the movement segment is based on changes in (at least one) hand movements. Also, if two hands produced independent signs simultaneously (e.g., *holding* with left hand and pointing to *the bird* with right hand, see Fig. 2-5, MS2) it was annotated as a single movement segment. Additionally, marked non-manual articulators (change in torso position, change in head position, facial expression, eye gaze direction) in each movement segment were annotated.

For example, in Figure 2-5, a signer encoded a stimulus from *Information density level 4* with referents being a dog and a bird. The signer first introduced a dog by means of the lexical sign *dog* in MS1. Then, in MS2 the signer pointed to himself (i.e., signaling assuming the role of the dog) and simultaneously encoded the action of holding through constructed action (CA). He then maintained the action of holding throughout the production while introducing the bird by first pointing to it (MS3) and then producing the lexical sign *bird* (MS4). Note that in MS3 head and eye gaze direction was referential as it often occurs together with index signs (Engberg-Pedersen, 2003). Accordingly, change in these non-manual parameters did not indicate marking of the referent but instead reinforced the index sign to the other referent being held. In the last movement segment, while maintaining the holding action, the signer encoded the bird *pecking* through the respective lexical sign (MS5). In MS5, the dog was marked via the head of the signer as receiver of the action of the bird. The bird, instead, was marked through the mapping of the bird's mouth onto the mouth of the signer.

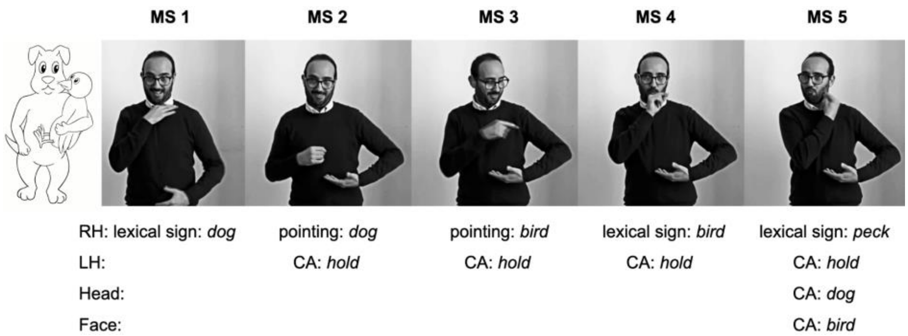


Figure 2-5. Example of the segmentation of the movement segments of a single stimulus (*Information density level 4*, referent pair: dog and bird): 5 movement segments in total.

It was crucial for us to start with movement segments as the base level, given that we then could proceed to unpack each movement segment in regard to whether it contained kinematic simultaneity and how many information units were encoded simultaneously in each movement segment. As such, movement segments represented the sequential nature of encoding. Once we had segmented data on the linear scale (sequentiality of movement segments) we then proceeded to assess whether these segments contained simultaneity (kinematic simultaneity and information density of simultaneity).

We excluded all movement segments that were clear disfluencies or mistakes after which signers corrected themselves. Also, given that we focused on how signers encoded 5 information units that we manipulated in different levels, we excluded additional movement segments that added extra information that was not the focus of our study (e.g., size or shape of the referents, movement segments encoding only eye gaze direction of the referents).

2.3.5.2. Kinematic simultaneity

We then coded whether articulators (manual and non-manual: left hand, right hand, torso, head, eye gaze, facial expression) were used simultaneously to encode different information units (referent 1, referent 2, static action, dynamic action 1, dynamic action 2) in each movement segment. If more than one articulator was used to encode different information units, the movement segment was coded as "simultaneous". Accordingly, this coding showed how many movement segments in each production contained kinematic simultaneity.

2.3.5.3. Information density of simultaneity

We then counted how many information units were encoded simultaneously within each movement segment. In the design, we constructed stimuli to focus on the following information units: referent 1, referent 2, static action, dynamic action of referent 1, dynamic action of referent 2. This added up to 5 information units that had to be encoded in our design, the specific amount of which was dependent on the information density level of the stimuli. In each movement segment, we counted how many information units of interest were simultaneously and explicitly available to the interlocutor. Thus, implicit referents (e.g., referent 2 implicitly available by the form of holding action, Figure 2-6) were not counted.

When encoding action by hand, referent had to be marked by at least one non-manual marker (eye gaze, facial expression, head, torso) in order to be counted as encoded (e.g., Figures 2-6, 2-7, 2-8). In cases where actions of both referents were encoded simultaneously, the second referent was only coded if it was explicitly marked by eye gaze, head, or facial expression (e.g., Figure 2-8). If not, we coded that only the action of referent 2 was encoded (e.g., Figure 2-7).

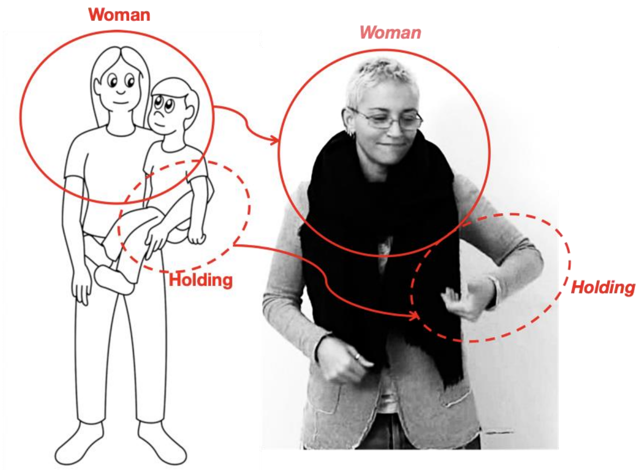


Figure 2-6. Movement segment of simultaneous encoding of *2* information units: ref. 1 - woman (encoded through head, facial expression and eye gaze) and the static action of holding (the left hand of the signer) (*Information density level 2*).

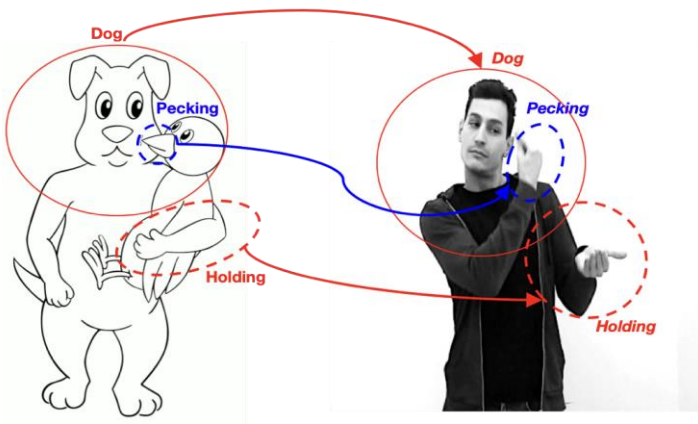


Figure 2-7. Movement segment of simultaneous encoding of 3 information units: ref.1- dog (encoded through torso, head, eye gaze and facial expression), static action of holding (left hand) and dynamic action of ref.2 - pecking (lexical sign *peck* signed by the right hand) (*Information density level 4*).

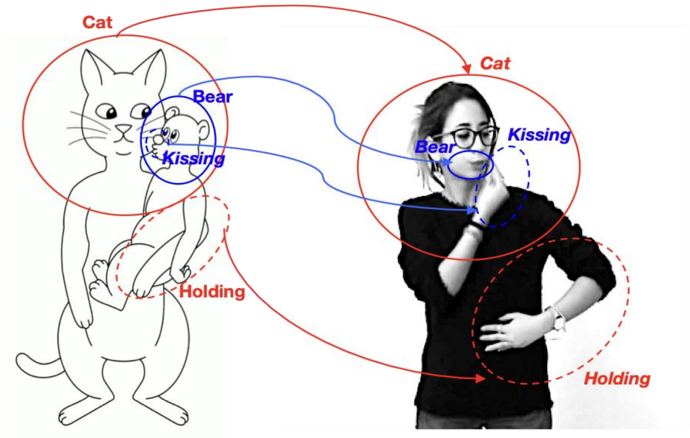


Figure 2-8. Movement segment of simultaneous encoding of 4 information units: ref.1 - cat (encoded through torso, head and eye gaze) and static action of holding (left hand), referent 2 (encoded through mouth) and dynamic action of ref.2 - kissing (lexical sign of *kiss* signed by the right hand) (*Information density level 4*).

2.3.6. Reliability

All data was initially coded by the first author of the study. All coded data was double-checked by a deaf researcher, who is a native signer of LIS. Another native signer of LIS independently coded 20% of the data.

Agreement between coders was almost perfect – 98.53% for gross-level segmentation of movement segments. Agreement assessment was done based on Kita et al (1997, p. 10): "The gross-level segmentation is recognition of a stretch of movement with a certain directionality [..] as a phase, regardless of the exact location of the boundaries and the identification of the phase type [..]". Out of 885 annotated movement segments coders agreed on 872 movement segments. We then derived a reliability statistic by assessing the total number of movement segments per stimuli. This was very strong as revealed by *Cohens* $\kappa = 0.94$.

Reliability of simultaneous use of multiple articulators (manual and non-manual) in each movement segment (*Cohen's* $\kappa = 0.94$) as well as reliability regarding the number of simultaneously encoded information units in each movement segment (*Cohens* $\kappa = 0.95$) was very strong.

2.3.7. Analyses

We analyzed data in R by using generalized mixed models (package *lme4*, Bates et al., 2015). The significance of the mixed model was derived from model comparisons starting with the baseline model that includes random effects and significant confounding factors. Significance levels were derived by using Satterthwaite's method provided by R package *LmerTest* (Kuznetsova et al., 2017). We use this method as it is proven to be the most conservative in regard to Type I errors, i.e., false positives (Luke, 2017).

We used the method of mixed effects models to test the effect of the *Information density level* on the *length of the production*, *kinematic simultaneity,* and *density of simultaneity* used. Mixed effects models make it possible to examine not only the fixed effects, but also include random effects of the individual trials and participants. Moreover, mixed effects models allow modeling not only of random intercepts but also random slopes and thus can account for even more fine-grained individual variation that might have an influence on the outcome of the analyses.

In our study, the following random effects were considered for the model: *trial (stimulus sample)*, *participant and referent pair*. By including a random intercept (i.e., random effect) for the stimulus sample and referent pair we account for possible variability that some specific stimuli or specific referent pair might be generally more powerful in eliciting simultaneity from the participants than others. Also, some participants might, in general, be more prone to use simultaneity than other participants, thus a random intercept for participant was used. It is also possible that the effect of the information density level is stronger for some participants than for others. Thus, in order to account for this aspect, we also considered the random slope of *Information density level by participant*. Accordingly, the individual differences of participants in regard to how sensitive they are to the predictor variable could be controlled for. Furthermore, we ran a series of models to account for possible confounding fixed factors, e.g., *gender, age, age of LIS acquisition, handedness.* The final baseline model was determined based on the best fit as revealed by ANOVA tests or alternatively on the maximal random effects structure that converged in the model (Barr et al., 2013). All stimuli, data and analyses scripts are available online (https://osf.io/mwg4v/).

2.4. Results

In the present experiment, we tested whether participants varied in the amount of kinematic simultaneity and information density of simultaneity, based on the information density level of the event they had to encode. The results section is organized as follows: first, we assess the differences in regard to the length of the production (number of movement segments used) in each information density level; next, we test whether kinematic simultaneity (2 or more articulators used simultaneously to encode different information) increases with the increase of the information density level; finally, we explore in which levels the most informationally dense simultaneous constructions (i.e., number of simultaneously encoded information units in a single movement segment) are used. We hypothesized that participants will increase kinematic simultaneity to encode informationally denser messages and that as the events that have to be encoded get denser so will the information density of simultaneity. The results are based on 23 participants describing 30 items depicting perceptually simultaneous events varying in their information density. We excluded 12 out of a total of 690 trials in which signers produced an incomplete description (e.g., omitted one or more information units). In order to be more conservative in regard to natural production of the participants, we did not consider corrected descriptions when prompted by the experimenter. Accordingly, the results are based on 678 trials in total.

2.4.1. Length of encoding

In Figure 2-9 we present raw means of the total number of movement segments (MS) per stimuli used in each information density level. The participants used on average 3.72 MS (SD=1.04) to encode 2 information units in level 1; 4.33 MS (SD=0.73) on average to encode 3 information units in level 2; 5.60 MS (SD=1.18) on average to encode 4 information units in level 3; 5.93 MS (SD=1.30) on average to encode 4 information units in level 4; and 7.68 MS (SD=1.64) on average to encode 5 information units in level 5. The statistical analysis was based on 678 data points (experimental trials/stimuli).

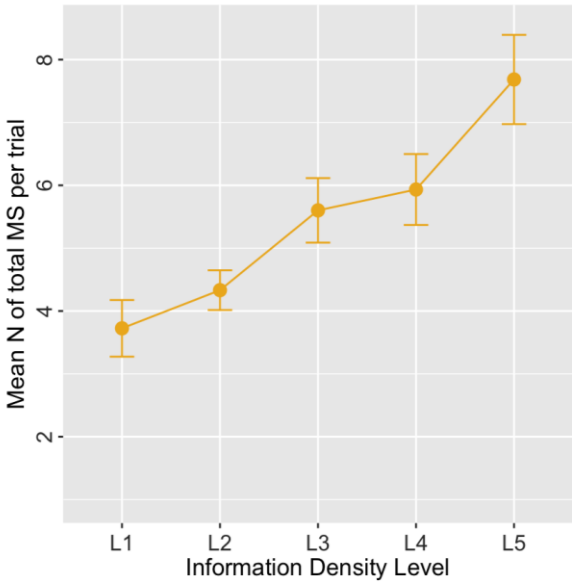


Figure 2-9. Mean of total number of movement segments used to encode each stimulus. Error bars indicate 95% CI of observations grouped within participants.

A generalized mixed effects model was fit to assess the fixed effect of *Information density level* (coded as categorical variable with 5 levels) on *the length of encoding* quantified as a count of the total number of movement segments used per stimuli (family = *poisson*). The effect of *Information density level* was compared to the baseline model which included random effect of *participant* and random effect of *trial*. We ran a series of models to test possible confounding factors - *gender, age, age of LIS acquisition, handedness and referent pair*. None of the factors was significant. In the remainder of the study we do not mention these factors unless they proved to be significant. There was a significant main effect of *Information density level* ($\chi 2(4) = 83.47$, $p < .001$, see Table 2-1 for a summary of the model).

The primary model was releveled in order to attain hierarchical contrasts between the levels. Pairwise comparisons revealed a significant gradual increase in the length of the production as the information that had to be encoded increased (level 1 vs. level 2: β=0.15, SE=0.06, CI[0.03, 0.27], *z*= 2.47, *p*=.01; level 2 vs. level 3: β=0.26, SE=0.05, CI [0.15, 0.37], *z*= 4.72, *p*<.001; level 4 vs. level 5: β=0.27, SE=0.05, CI[0.17, 0.36], *z*= 5.62, *p*<.001) except for levels 3 and 4 were comparable (β=0.05, SE=0.05, CI[-0.04, 0.15], *z*= 1.07, *p*=.28).

Random effects	**Variance**		**SD**			
Trial	0.00[10]		0.00			
Participant	0.02		0.15			
Number of obs: 678	Groups: Trial=30,		Participant=23			
Fixed effects	**95% CI**		**β**	**SE**	***z***	***p***
	Lower b.	Upper b.				
(Intercept)	1.20	1.41	1 .30	0.05	23.94	**<.001**
Level 2	0.03	0.27	0.15	0.06	2.47	**.01**
Level 3	0.30	0.52	0.41	0.06	7.18	**<.001**
Level 4	0.35	0.57	0.46	0.06	8.18	**<.001**
Level 5	0.62	0.83	0.73	0.05	13.47	**<.001**

Table 2-1. Best fit model in a logit scale (model fit by maximum likelihood, Laplace Approximation) regarding use of total number of MS per experimental stimuli. Contrasts reflect pairwise comparison between level 1 and all other levels.

2.4.2. Kinematic simultaneity

Considering the differences in the length of the production, analysis was based on proportions of movement segments (MS) with kinematically simultaneous articulators (i.e., two or more articulators used in a single movement segment) out of the total number of movement segments used per stimulus (Figure 2-10). The movement segments expressed with kinematically simultaneous articulators were scarcely used (M=0.05, SD =0.07) in the least dense information level (level 1). In level 2 almost half of the MS (M=0.46, SD = 0.16) contained kinematic simultaneity. Kinematic simultaneity increased further as the information density of the event that had to be

[10] Zero variance of the random effect of trial is driven by inclusion of the fixed effect of *Information density level*, which accounts for all variance detected in random effect of trial in the baseline model. Given that inclusion of trial is based on the initial design of the study and the results do not change if this random effect is left out, we keep it in the primary model. Controls of random effect of trial can be found in supporting material. This consideration applies to all consecutive analyses.

encoded increased: level 3 (M=0.57, SD =0.12), level 4 (M=0.60, SD =0.10), and level 5 (M=0.67, SD =0.10). The statistical analysis was based on 678 data points (experimental trials/stimuli).

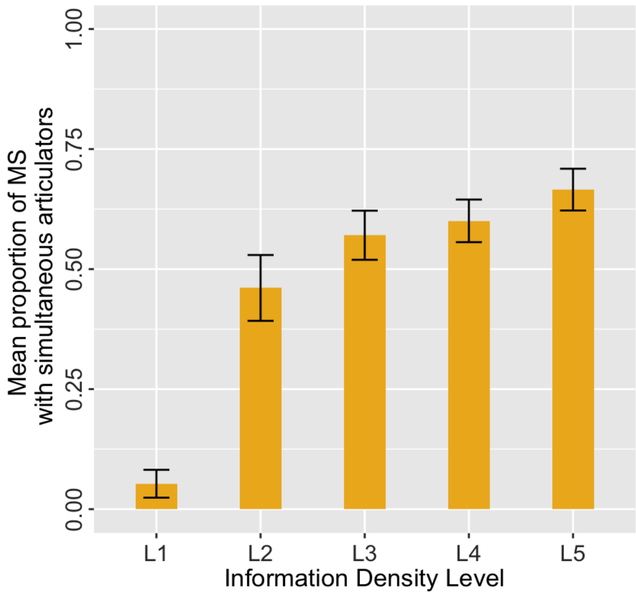


Figure 2-10. Raw mean proportions of kinematically simultaneous movement segments out of total number of movement segments used per stimuli. Error bars indicate 95% CI of observations grouped within participants.

A generalized mixed effects model was fit to assess the fixed effect of *Information density level* on the amount of *kinematic simultaneity* used, quantified as total number of MS containing simultaneity versus total number of MS used to encode each trial (family = *binomial*). The effect of *Information density level* was compared to the baseline model, which included random effect of *participant* and random effect of *trial.* The main effect was highly significant ($\chi2(4) = 110.16$, $p < .001$, see Table 2-2 for a summary of the model).

The primary model was releveled in order to attain hierarchical contrasts between the levels. Pairwise comparisons revealed that there was a significant gradual increase in use of MS with kinematically simultaneity (level 1 vs. level 2: β=2.61, SE=0.20, CI[2.22, 3.00], z= 13.19, $p < .001$; level 2 vs level 3: β=0.47, SE=0.11, CI[0.25, 0.69], z= 4.15, p < .001; level 4 vs. level 5: β=0.33, SE=0.10, CI[0.13, 0.52], z= 3.26, p= .001) except for levels 3 and 4 which were comparable (β=0.08, SE=0.1, CI[-0.12, 0.29], z= 0.77, p = .44).

Random effects	**Variance**		**SD**			
Trial	0.00		0.00			
Participant	0.15		0.39			
Number of obs: 678	Groups: Trial=30,		Participant=23			
Fixed effects	**95% CI**		**β**	**SE**	***z***	***p***
	Lower b.	Upper b.				
(Intercept)	-3.12	-2.35	-2.74	0.20	-13.91	**<.001**
Level 2	2.22	3.00	2.61	0.20	13.19	**<.001**
Level 3	2.70	3.46	3.08	0.19	15.87	**<.001**
Level 4	2.78	3.54	3.16	0.19	16.30	**<.001**
Level 5	3.11	3.86	3.49	0.19	18.17	**<.001**

Table 2-2. Best fit model in a logit scale (model fit by maximum likelihood, Laplace Approximation) regarding the proportion of kinematically simultaneous MS. Contrasts reflect pairwise comparison between level 1 and all other levels.

2.4.3. Information density of simultaneity

Overall, there were 3697 movement segments used, out of which there were 1748 MS with one information unit, 1225 MS with two information units, 622 MS with three information units, and 102 MS with four information units. Accordingly, the results revealed that a maximum of four information units were encoded simultaneously in our data set (Figure 2-11). These were found only in *Information density levels 4 and 5* (except one instance of four information units in level 3). MS with two simultaneously encoded information units were found in all levels. Note that in level 1 all kinematically simultaneous movement segments contained exclusively 2 information units. The statistical analysis was based on 3697 data points (movement segments).

In order to test whether information density of simultaneity increased based on the increase of information density level, we ran a generalized mixed effects model where the fixed effect was *Information density level* and the outcome variable was the *Information density of simultaneity* quantified as the count of total number of information units encoded in a single MS (family = *poisson*). The effect of

Information density level was compared to the baseline model, which included random effect of *participant* and random effect of *trial.* The main effect was highly significant ($\chi2(4) = 88.74$, $p < .001$, see Table 2-3 for a summary of the model).

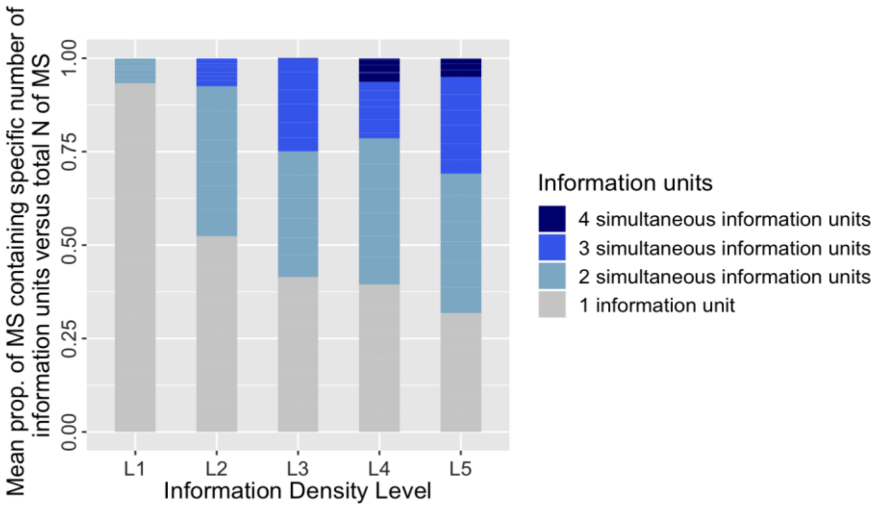


Figure 2-11. Raw mean proportions of MS with 1, 2, 3, and 4 simultaneous information units out of the total number of movement segments.

The primary model was releveled in order to attain hierarchical contrasts between the levels. Pairwise comparisons revealed that as information density level increased the movement segments used to encode these levels became denser in regard to the number of information units that they contained (level 1 vs. level 2: β=0.38, SE=0.05, CI[0.27, 0.48], *z*= 6.92, *p* < .001; level 2 vs level 3: β=0.17, SE=0.04, CI[0.09, 0.25], *z*= 4.00, *p* < .001; level 4 vs. level 5: β=0.08, SE=0.03, CI[0.01, 0.14], *z*= 2.29, *p* = .02) except for levels 3 and 4 which were comparable (β=0.02, SE=0.04, CI[-0.05, 0.1], *z*= 0.68, *p* = .50). It appears that the lack of difference in increase of density of simultaneously encoded information is driven by signers using more MS with 2 information units in level 4 than in level 3. As a result, the incremental use of MS with 4 information units is balanced out at the expense of higher use of MS with 2 information units.

Random effects	Variance		SD			
Trial	0.00		0.00			
Participant	0.001		0.03			
Number of obs:3697	Groups: Trial=30,		Participant=23			
Fixed effects	**95% CI**		**β**	**SE**	***z***	***p***
	Lower b.	Upper b.				
(Intercept)	-0.02	0.15	0.06	0.04	1.43	**.15**
Level 2	0.27	0.48	0.38	0.05	6.92	**<.001**
Level 3	0.45	0.64	0.55	0.05	10.84	**<.001**
Level 4	0.47	0.67	0.57	0.05	11.43	**<.001**
Level 5	0.55	0.74	0.65	0.05	13.49	**<.001**

Table 2-3. Best fit model in a logit scale (model fit by maximum likelihood, Laplace Approximation) regarding use of MS with increasing information density (N of information units MS contains). Contrasts reflect pairwise comparison between level 1 and all other levels.

2.5. Discussion

In the present study, we hypothesized that as the amount of information to be communicated increases, so will the use of kinematic simultaneity and information density of simultaneity. Our data confirmed that not only the length of production increased but also the use of kinematic simultaneity. Furthermore, as the information to be communicated became denser, so did the simultaneously encoded information in the movement segments (MS). These findings indicate that, at least in LIS, signers use simultaneous information encoding to achieve efficient communication.

Given that not only simultaneity but also the length of productions increased following the increase in information demands indicates that simultaneity was not simply used to reduce the number of movement segments (which might seem a straightforward prediction), but it was used for simultaneous encoding of multiple related information units. The increase in both length and simultaneity might lie in the contrasting needs to be both informative and brief (Grice, 1975; Perniss, 2007b). Presenting an entire utterance in a single simultaneous construction right away might lead to ambiguity, while stacking one or multiple units on existent information may accommodate for informative and temporal needs. Alternatively, presence of the increase in linearity might also be linked to motoric constraints (e.g., producing two actions simultaneously requires more planning (Oliveira & Ivry, 2008)) or absence of time constraints (signers were not under pressure to sign as fast as possible). However, systematic comparisons with spoken languages are needed, to fully understand the relation between simultaneity and length of production in sign languages. So far comparisons of information encoding in spoken and sign languages have shown that in cases in which the amount of information is kept constant fewer consecutive linguistic units are needed in sign languages compared to spoken languages (Bellugi & Fischer, 1972). However, further studies are needed to evaluate whether this would also be true in cases in which the density of information to be conveyed grows, such as in the present study. We expect that there would be an increase in linguistic units in sign and spoken languages, but that the former would be less than the latter. A systematic comparison assessing this prediction should be an endeavor for further research.

In regard to information density of simultaneity, MS containing two information units were used in all levels, use of MS with three information units were used in levels 3, 4 and 5 considered in our study; MS with four information units were used only in levels 4 and 5 (with exception of one instance in level 3). Interestingly we never found five information units encoded simultaneously, although, at least for half of the stimuli in levels 4 and 5, this would be possible given that the action performed by the referent 2 was done with the head/mouth (i.e., beak, tongue, lips). One could have expected that if direct mapping is the aim of the encoding, we would find use of non-manual articulators to encode the third action as well. However, data suggests that this is not the case and, at least in LIS, actions are encoded also using the hands. An alternative explanation, however, might stem from cognitive constraints. Research shows that there is a limit on visual memory capacity which appears to be exactly four elements/objects (Cowan, 2001, 2010; Irwin & Andrews, 1996; Luck & Vogel, 1997; Sperling, 1960; Vogel et al., 2001). This is also in line with findings from Napoli & Sutton-Spence, (2010) who find a maximum of 4 propositions in ASL. However, research on visual working memory capacity disregards the fact that diagrammatic iconicity may bind multiple elements into single representation. In this respect, it is not clear whether each information unit encoded simultaneously in sign language should be treated as an independent element, and thus the limit of maximum 4 elements would apply, or instead as a feature of a single more complex element.

Additionally, we found that in levels 3 and 4 in which we maintained the same number of information units but manipulated the complexity of the event (i.e., one agent versus two agents), the quantitative increase in use and density of simultaneity was not significant. Signers encoded these two events using the same proportion of MS containing kinematic simultaneity by manipulating the density of the simultaneously encoded information in a qualitative manner (i.e., using more MS containing 2 and 4 simultaneous information units in level 4 in comparison to level 3). Although assessment of the specific information encoded in an individual MS was not the goal of our study, we speculate that it is conceptually easier to encode a single referent and its actions simultaneously as opposed to simultaneous encoding of two referents and their actions, as the latter requires splitting the body in two conceptually distinct entities through body partitioning. The results also indicate that it is possible that for some signers the ease of encoding might influence their striving for maximum simultaneity. Namely, instead of always encoding four information units in the same MS, signers could instead encode two separate MS each encoding one referent and its action. Accordingly, our choice to add level 4 that increases complexity while maintaining the same number of information units as in level 3 seems justified as it provides some preliminary insights into how the use of simultaneity is affected by the complexity of the event while information density is held constant. In line with the results discussed in the previous paragraph, in the future it would be necessary to explore whether simultaneous encoding is constrained by cognitive demands and/or by articulatory effort.

2.5.1. Simultaneity over linearity for efficient communication in sign languages

Our results show that as the information that had to be encoded increased, signers increased the amount and density of simultaneously encoded information, which was only possible due to the bounding property of diagrammatic iconicity. In other words, diagrammatic iconicity allowed signers to intertwine additional information with the schematization resulting in a single representation. If anything, without diagrammatic iconicity, simultaneous encoding of multiple information units, with no iconic relation connecting them, would hinder efficient communication. We argue that signers in our study strived for a more direct mapping of the meaning and linguistic encoding in order to boost more efficient representation formation.

This interpretation appears to also be in line with Christiansen & Chater (2016) who argue that given the constraints of perceptuo-motor processing and memory "the language system engages in eager processing" (p.5) in order to create higher-level linguistic representation as fast as possible. Namely, low-level information is passed onto higher representation levels, e.g., sounds → words → discourse, to form a single chunk of representation which can be retained more efficiently. Note that for spoken languages the low-level chunks (i.e., sounds) are passed onto higher levels in a linear manner. For sign languages, on the other hand, chunking can occur both linearly, i.e., chunking one sign after another, and simultaneously due to diagrammatic iconicity, that relates multiple meaning elements that are produced with different articulators. As a result, use of diagrammatic iconicity would boost the chunking process, considering that a higher-level representation could be constructed in a direct relation to other information in a simultaneous manner as opposed sequentially. Signers use mental imagery generation for language production and comprehension (Emmorey, 1993; Emmorey et al., 1993; Emmorey & Kosslyn, 1996) especially if spatial relations are involved (Emmorey, 1995). Thus, the advantage of the possibility to exploit diagrammatic iconicity to bind different information units into single representation might be at play in encoding and decoding, as it can be used to adhere to the mental imagery of the signer and interlocutor. As a result, use of diagrammatic iconicity for simultaneous encoding of information would boost the chunking process, considering that a higher-level representation is available right away as opposed sequentially. Indeed, diagrammatic iconicity has been shown to also aid conceptual processing in non-signers (Louwerse & Jeuniaux, 2010; Zwaan et al., 2002; Zwaan & Yaxley, 2003), indicating that it can be used as a tool in more efficient representation forming involving multiple elements. For example, Zwaan & Yaxley (2003) showed that simultaneously presented words in reverse-iconic relation (e.g., the word *basement* above the word *attic)* resulted in slower semantic-relatedness judgments than words presented in iconic relation (the word *attic* above the word *basement*). We speculate that exactly simultaneous encoding in sign languages is an efficient strategy to chunk information in higher level representations. Vinson et al. (2015) argue for iconicity as a vehicle for a faster path between meaning and form in production and comprehension on a single lexical sign level. We predict that this finding would also extend to iconic simultaneous constructions. Testing this hypothesis, however, would be an endeavor for future research.

The specific sensory constraints present in language production might put specific pressures on how language is organized leading to the exploitation of simultaneity over linearity in sign languages (Supalla, 1991). Signers are adept at dealing efficiently with the integration of simultaneous information. Especially considering signs' multilinear nature and use of space for grammatical encoding that has to be processed regularly during language use (Wilson & Emmorey, 1997). Indeed, Capirci et al. (1998) show how even minimal, but constant, exposure to a sign language can boost visual-spatial cognition and spatial memory in hearing children. Therefore, it is not surprising to suggest that when it comes to proficient sign language users, they may take advantage of simultaneous information encoding in order to cluster related meanings closer together to increase communicative efficiency. As a result, it would be less cognitively demanding for signers to encode and decode information presented simultaneously as opposed to consecutively. Research suggests that when visual working memory is involved it benefits when simultaneity can be taken advantage of (Allen et al., 2006; Frick, 1985; Wilson & Emmorey, 1997; Woodman et al., 2003). The same is not true when processing of auditory information is involved, which seems to be bound to a persistent unidirectional effect, i.e., auditory items are recalled in the same sequence as the one in which they have been presented (McFarland & Kellas, 1974; Penney, 1989). Indeed, serial recall is particularly hard for signers in comparison to speakers as indicated by studies documenting shorter memory span in signers (Bavelier et al., 2006; Boutla et al., 2004; Geraci et al., 2008; although see also Rudner & Rönnberg, 2008; Wilson & Emmorey, 2006). We argue that the finding that signers increase use of simultaneity as information to be conveyed becomes informationally denser, suggests that simultaneity is used strategically to cluster related meanings as close together as possible to boost efficient communication through easing processing. For example, when a signer introduces a referent and subsequently becomes the referent in order to encode the action of holding, the referent does not have to be maintained solely in working memory as it is conceptually present through the signer's body. Furthermore, instead of providing each information element one at the time, additional information can be integrated into the iconic diagram to form a more complete representation. Accordingly, to some extent, simultaneous information encoding could function as cognitive offloading and lighten processing (Risko & Gilbert, 2016) considering that some information can be externalized and maintained as new information is introduced. Indeed, Napoli & Sutton-Spence (2010, p.675) also suggest that *connectedness* among simultaneous units is a potential candidate for reducing cognitive load. Whether simultaneous information encoding can indeed lighten cognitive load in sign languages may be tested in the future on both behavioral and neurological levels.

Although we stressed that spoken languages have been described as strictly linear, use of the whole body or of individual bodily parts would also allow speakers to employ bodily articulators to encode distinct information simultaneously (Kendon, 2014). However, it was not our intention to directly compare simultaneous information encoding in sign languages to spoken languages. Accordingly, discussion on multimodality in spoken languages was not elaborated on. While it is expected that signers are more sophisticated in using their body for information encoding, it would be interesting to see whether speakers also take advantage of the visual modality for efficient communication through simultaneity (e.g., use of complementary gesture with speech). Previous research indicates that, for the purpose of task-solving, use of gesture with speech can lighten working memory load (Cartmill et al., 2012; Cook et al., 2012; Goldin-Meadow & Beilock, 2010).

Furthermore, future research on grammatical structure would be needed to understand how use of simultaneity interacts with affordances and the constraints of grammatical structure of a specific language. In particular, it would be interesting to explore which information exactly is clustered together and how, in both spoken (speech + gesture) and sign languages. In this vein, it would be important to investigate specific linguistic strategies used for simultaneous encoding of information. We described in brief how different linguistic strategies and their iconic affordances allow simultaneity in sign languages, but a systematic assessment of specific linguistic strategy use and their combinations would be necessary to elucidate how exactly communicative efficiency through simultaneity is achieved. The study on the choice of specific linguistic strategies used for communicative efficiency by LIS signers is reported in **Chapter 3**.

Finally, it would be crucial to assess how hearing non-signers can use their body to transmit information at different information density levels when no linguistic encoding is possible (i.e., silent gesture paradigm). Comparisons between signers and silent gesturers can elucidate how much of the iconic structures and simultaneity we find here are due to the visual modality affordances of using multiple articulators and iconicity and how much due to skillful use of linguistic resources. It is highly possible that use of more simultaneous and iconic constructions requires complex linguistic tools and as such simultaneity and iconicity might constitute linguistic properties that emerge into linguistic structure for communicative efficiency. A study on these topics where we ask Italian speakers to express the same stimuli using silent gestures and compare those productions to those of LIS signers is reported in **Chapter 4**.

2.5.2. Limitations of the study

There are some limitations to this study that we consider important to stress. First, we were interested in assessing how many information units can be encoded simultaneously in LIS. In our design, we could test simultaneous encoding of two animate referents and their actions. Accordingly, we cannot generalize our findings to overall use of simultaneity in sign languages. Nevertheless, this is the first systematic study to quantify simultaneous encoding of information of events involving animate referents and to assess whether this property is used to achieve communicative efficiency in LIS. Future research should show whether our findings generalize to other sign languages and other types of events as well.

Second, based on our results, we find that out of a maximum of five information units that we focused on, four units could be encoded simultaneously. Even though we criticized Napoli & Sutton-Spence (2010) for not accounting for the overall simultaneity that sign languages are able to achieve, we may be seen as encountering a similar problem. In fact, even in our design there was more information available to the interlocutor that could be simultaneously encoded by the participants than what we quantified in our data, (e.g., eye gaze direction, action location, and size of the referents). There is yet another information unit that we disregarded, namely, the availability of the implicit referent. Dudis (2004) notes that signers can render non-present referents available to the interlocutor via conceptual integration. For example, in instances of signers encoding referent 1 and the holding action, the non-present but yet conceptually available referent is the one that is being held, i.e., implicit referent (Fig. 2-6). The same applies for instances where signers encode referent 2 and dynamic action 2 in a single movement segment (see Fig. 3-7). The action in this case is performed on a non-present but yet conceptually available referent 1. In the future, it would be important to account for all the information that is available to the interlocutor in a single movement segment and how it is increased. However, this would require different stimuli and more fine-grained analyses specifically targeting this issue in order to ensure an adequate study design. Another interesting question that arises is whether there are differences in cognitive effort when integrating implicitly vs. explicitly encoded information.

2.6. Conclusion

An event in the world has to travel a long way to reach its encoding in language. Just as in spoken languages, also in sign languages signers face the linearization problem (Levelt, 1980). Namely, an event that occurs simultaneously in the world has to be split in language to be organized on a temporal scale. Accordingly, the goal of clustering related information closer together in order to render communication more efficient stands also for sign languages. While striving to cluster related information closer together can be considered a general trait of language, due to temporal processing constraints (Christiansen & Chater, 2016), how this is achieved is, to some extent, depends on the linguistic modality. We hypothesized that in sign languages (i.e., LIS in this specific case) clustering related information could be achieved by exploiting diagrammatic iconicity to encode information units of the same event simultaneously. Signers do not limit their clustering of related information consecutively as in spoken languages (Lu et al., 2016), but they also encode information simultaneously, minimizing dependency distances. We attempted to quantify how simultaneous encoding of information is used by signers of LIS when encoding events with various information densities. We found that kinematic simultaneity was used more and information density of simultaneity increased when it was necessary to provide more information.

The core property of iconicity to depict the world as opposed to arbitrariness that describes it "is one of the central affordances of human language" (Dingemanse, 2018, p. 19). It reinforces the link between the form and the meaning and allows more direct information transmission (Perniss & Vigliocco, 2014; Vinson et al., 2015). Signers employ iconicity to represent the information present in events as it is available in the real word – simultaneously – and as such they are more truthful to the facts they are referring to. As a result, conceptual representation can be formed faster. We conclude that iconicity in sign languages should be seen as an advantage, as it allows more efficient communication through simultaneity.

2.7. Acknowledgments

We wish to thank Barbara Pennacchi for drawing the stimuli for the experiment and assistance with data collection. We wish to thank Alessio Di Renzo, Luca La Mano, Tommaso Lucioli, and Alessandra Ricci for assistance during data collection and coding. Furthermore, we wish to thank Sean Roberts and Roberta Rocca for invaluable help with statistical analyses, and Laura Sparaci for proofreading the manuscript.

Chapter 3

Using depiction for efficient communication in LIS (Italian Sign Language)

This chapter is based on:

Slonimska, A., Özyürek, A., & Capirci, O. (2021). Using depiction for efficient communication in LIS (Italian Sign Language). *Language and Cognition, 13*(3), 367-396. doi:10.1017/langcog.2021.7

Abstract

Meanings communicated with depictions constitute an integral part of how speakers and signers actually use language (Clark, 2016). Recent studies have argued that, in sign languages, depicting strategy like constructed action (CA), in which a signer enacts the referent, is used for referential purposes in narratives. Here, we tested the referential function of CA in a more controlled experimental setting and outside narrative context. Given the iconic properties of CA we hypothesized that this strategy could be used for efficient communication. Thus, we asked if use of CA increased with the increase in the information required to be communicated. Twenty-three deaf signers of LIS described unconnected images of perceptually simultaneous events, which varied in the amount of information represented, to another player in a *director–matcher* game. Results revealed that participants used CA independently and in combination with other linguistic strategies to communicate core information about the images and also increased the use of CA as images became informationally denser. The findings show that iconic features of CA can be used for referential function in addition to its depictive function outside narrative context and to achieve communicative efficiency.

3.1. Introduction

When we look at the natural niche in which language occurs, that is, in face-to-face interactions, it is evident that not only are arbitrary and categorical properties used to express meaning (Hockett, 1978, 1960), but also meanings communicated with depictions through iconic representations. In fact, the distinction between the descriptive and depictive properties of language has been made for at least the last 15 years (Clark & Gerrig, 1990; Cormier, Smith, et al., 2015; Ferrara & Hodge, 2018; Holt, 2000; Liddell, 2003, among others). The definition of depiction, and its contrast with description, is well captured by Clark (2016, p. 342): "To describe something is to tell others about its properties—to represent it categorically. […] To depict something, however, is to show others what it looks or sounds or feels like."

In recent years there has been an ever-growing interest with regard to depiction as an integral part of linguistic structure in both sign and spoken languages (e.g., Clark, 2016, 2019; Dingemanse, 2018; Dingemanse et al., 2015; Dudis, 2004, 2011; Ferrara & Halvorsen, 2017; Kendon, 2014; Müller, 2018). In the present study, we focus on so-called *depiction* and show that in sign languages it can also actually be used for descriptive or referential purposes, and in particular to attain communicative efficiency. Communicative efficiency can be described as a fundamental property that shapes the structure of languages "to facilitate easy, rapid, and robust communication" (Gibson et al., 2019, p. 389).

In a previous study, Slonimska et al. (2020) investigated whether signers use simultaneity, a property afforded by use of multiple articulators and iconicity, for achieving communicative efficiency. Namely, authors assessed whether signers increased their use of simultaneous constructions to encode information as a function of an increase in the amount of information that needed to be communicated. In this study, the amount of information signers needed to encode was experimentally manipulated, in a non-narrative context, and as the amount of information that needed to be communicated by the signers increased, so did the use of simultaneous constructions. The present study uses the same data and experimental manipulation of communicative efficiency as in Slonimska et al. (2020) and further investigates whether greater need for communicative efficiency also results in greater use of depiction, something that has not been investigated in the previous or any other study.

In sign languages, depiction plays a prominent role, given the rich iconic potential of the visual modality in which these languages are realized. Depictions in sign languages can be grouped into two types: depictions from an observer perspective and depictions from a character perspective (Perniss, 2007a, but for different terminology see Kurz et al., 2019). Manual depictions from the observer perspective are called *depicting constructions* (also called *classifier constructions, classifier predicates, polycomponential verbs*, and *polymorphemic verbs*; see Schembri, 2003). These constructions depict events in the signing space in front of the signer on a miniature scale. Non-manual depictions known as *mouth gestures* or *mouth actions* (Boyes-Braem & Sutton-Spence, 2001) can be used to provide adjectival or adverbial information in respect to the manual depicting constructions (Crasborn et al., 2008; Fontana, 2008). Depictions from the character perspective, in contrast, put the signer's body at the center of the production as the signer projects the referent directly onto their body and depicts the actions performed by the referent with corresponding body parts in life-sized scale (Cormier, Smith, et al., 2015; Kurz et al., 2019; Perniss, 2007a). Such a depicting strategy is called a *constructed action* (Metzger, 1995; Tannen, 1989), and it is the focus of the present study.

Recent studies have argued that *constructed action (CA)* is used for referential purposes, including encoding the core meaning elements, i.e., argument and predicate (Cormier, Smith et al., 2013; Ferrara & Johnston, 2014; Hodge & Ferrara, 2014; Hodge & Johnston, 2014; Jantunen, 2017; Pizzuto et al., 2006). However, most research on CA is embedded in a narrative context. Such a context might pose a problem in assessing the whole spectrum of the referential capacity of CA considering that a crucial factor in narration is the evaluative function (or emotive function, in the terms of Jakobson, 1960), which is used to enhance referential information (Labov & Waletsky, 1967). Accordingly, when looking at encodings of narratives it becomes practically impossible to tease apart whether CA is used because of its contribution to the evaluative function (i.e., making narration more vivid and entertaining through depiction) or for referential purposes (i.e., to encode the core meaning elements of the event), or a mix of the two. For example, research has shown that in narratives, the same content can be signed with or without CA, indicating that its use is not obligatory but rather can be a matter of "idiosyncratic preferences, storytelling experience, and sociolinguistic effects such as age and education" (Hodge & Ferrara, 2014, p. 388). Accordingly, in order to truly comprehend CA's referential capacity, it also has to be studied in contexts in which the necessity for referential function is unquestionable, as in contexts where information has to be communicated efficiently.

Slonimska et al. (2020) hypothesized that when signers are faced with increasing information encoding demands they might achieve communicative efficiency in a comparable way as spoken languages do, i.e., by minimizing dependency distances. Dependency distance minimization refers to a tendency of language users (studied only in spoken languages so far) to cluster semantically and syntactically related words closer together (Gibson et al., 2019; Temperley & Gildea, 2018). This strategy has been argued to lead to faster access to syntactic and semantic representation in production and comprehension (Hawkins, 2004), and thus to increase communicative efficiency (Gibson et al., 2019). Slonimska et al. (2020) were interested in exploring whether sign language users exploited multiple articulators and iconicity for encoding multiple information units simultaneously, considering that dependency distances could be reduced to the minimum in this way. Thus, they assessed whether signers increase the use of simultaneous constructions with the increase of the information that is required to be communicated. They found that this was indeed the case. For example, signers could encode information about the agent, patient, and their actions (e.g., a stimulus representing a cartoon image of a woman holding a boy and the boy pinching the cheek of the woman) in a single simultaneous construction as opposed to encoding each piece of information in a one-by-one fashion. Not only did the signers increase the encoding of information in a simultaneous as opposed to a strictly linear manner, but they also increased the density of the simultaneously encoded information. Information density of simultaneity was quantified as the number of simultaneously encoded semantic information units. While the aforementioned study provides evidence that signers use more simultaneous constructions when faced with increasing information demands, the linguistic strategies used and the role of depictions in achieving communicative efficiency still remain to be explored.

The aim of the present study is twofold. First, we aim to extend the assessment of the referential function of CA to a controlled experimental context through a study designed to elicit strictly referential information, thereby reducing to a minimum the need for the evaluative function. Second, we aim to assess whether CA is used to achieve efficient communication by way of an experimental design (used in Slonimska et al., 2020) in which signers are required to encode events with increasing information density (i.e., the number of semantic information units that need to be encoded). In such a setting a signer is expected to communicate in a way that encodes the message efficiently in terms of minimizing their own effort as well as making the message informative enough for the addressee (Gibson et al., 2019; Grice, 1975). As the information demands increase, the task of accommodating both of these aspects becomes harder. As a result, we expect that when signers are faced with increasing information encoding demands they will be likely to employ linguistic strategies which lead to efficient communication. Thus, if CA use increases as the amount of information to be communicated also increases, it would serve as a strong indicator that this strategy is used with referential purpose in order to achieve communicative efficiency.

3.1.1. Constructed action and types of iconicity

Constructed action (Metzger, 1995; Tannen, 1989), also known as *role shift* (Padden, 1986; Quer, 2011), *transfer of person* (Cuxac, 1999, 2000; Cuxac & Sallandre, 2007; Pizzuto et al., 2006; Volterra et al., 2022), and *enactment* (Ferrara & Johnston, 2014; Hodge & Johnston, 2014), is a depicting strategy attested in a plethora of sign languages (see Kurz et al., 2019) and is when the signer uses one or more bodily articulators, including hands, torso, head, eye gaze, and facial expressions, to directly map the referent to the signer's corresponding body part. Accordingly, the event depicted is represented as if it were from the perspective of the character involved in the event. Thus, the actions performed or feelings expressed by the referent are encoded by the signer depicting the actions and/or feelings with their own upper body.

Such depiction might sound quite familiar to non-signers considering that speakers also make use of a vast array of depictions, including *demonstrations* and character viewpoint gestures reminiscent of CA (Clark, 2016; Clark & Gerrig, 1990). Possibly for this reason, CA has been mostly regarded as exploiting only *imagistic iconicity,* i.e., resemblance between the form of the sign and its meaning (Cuxac, 1999, 2000; Perniss, 2007a; Taub, 2001), and thus as representing the referent and all its properties holistically (e.g., Ferrara & Johnston, 2014; Hodge & Ferrara, 2014; Jantunen, 2017). While some research does identify sub-elements out of which CA is actually constructed, it appears to be treated mainly as a degree of how intensely the referent depicted by CA is marked (Cormier, Smith, et al., 2015). For example, Cormier, Smith et al. (2015) propose that CA can vary in how intensely it marks a depicted character based on how many articulators are used in the construction. That is, different articulators can be used to varying degrees and thus CA can be considered as being overt, reduced, or subtle. Under this view, the signer chooses how strongly to mark the imagistic resemblance between the referent and the depiction. Cormier, Smith, et al. (2015) also mention the possibility of using a type of mixed CA (although it was not attested in their data), in which two or more characters can be encoded simultaneously. However, they also note that "the situations when [mixed CA] may be expected to occur are not well understood" (Cormier, Smith, et al., 2015, p. 192).

In the present study we argue that the use of a varying number of articulators during CA can not only be considered a stronger or weaker character marker but also a tool to encode different information by means of different articulators, and thus can be used for informative rather than intensifying purposes. For example, a signer who tilts their head upwards while depicting a person shaking hands does not only intensify the depiction of the character but also provides information in its own right, i.e., that the person is shorter than the person he or she is shaking hands with. Furthermore, this example also illustrates that both the articulators and their relation to each other provide information that is necessary for the decoding. In other words, we argue that CA possesses not only *imagistic iconicity* but also *diagrammatic iconicity*, i.e., the relation between the components of the sign or the construction representing the relation between the components of meaning (see Perniss, 2007a, for an overview of views on imagistic versus diagrammatic iconicity). If such a view is adopted, then the use of specific body articulators in CA does not necessarily function as a stronger marker of CA but instead serves to integrate multiple pieces of information about the event into a single representation more efficiently.

Given that the signer's body is central for CA, the articulators can be interpreted in a diagrammatic fashion – the information encoded by the hands and their relation to each other as well as the hands in relation to the information encoded by the body (Meir et al., 2007). For example, a signer can establish different diagrammatic relations by using diverse articulators: the signer can integrate information about space/direction with hand and torso movement (a woman pinching a child to her right with her right hand), and also add deictic information with the eye gaze direction and a referent's emotional state with a facial expression (a woman pinching a child while lovingly gazing at the child; Figure 3-1a). All those little details alter the interpretation of the depiction not only in an imagistic but also a diagrammatic fashion, since such alterations inevitably establish new relations between sub-components of the construction.

Furthermore, CA allows for the encoding of not only the same referent and its actions but also for the encoding of multiple referents and their relation to each other by depicting one referent and/or its actions with some articulators while encoding the other referent and/or its actions with other articulators (e.g., a child being pinched on the left cheek by a person taller than the child (a woman) on the left; Figure 3-1b). The strategy of splitting the body in order to encode different referents is known as *body partitioning* (Dudis, 2004), or *mixed CA type*, in the terms of Cormier, Smith, et al. (2015). Such constructions involve not only imagistic properties but also diagrammatic schematization of the event, which arguably makes CA an efficient strategy for encoding complex events involving multiple information elements (e.g., agent, patient, and action). Accordingly, CA can be viewed as not simply a more or less intense imagistic depiction of the referent but as a diagrammatic depiction in which multiple articulators are employed and the specific information they convey are interrelated and increase the informativeness of the message.

a b

Figure 3-1. Diagrammatic properties of CA when encoding relations between two referents and their interaction. a) CA depicting a woman pinching a child on her right (implicit) with her right hand while gazing lovingly at the child. The woman and her actions are mapped onto the body of the signer. The child is implicitly marked by direction of the action. b) CA depicting a child being pinched on the left cheek by a person taller than the child (the woman) on the left. The child is mapped onto the body of the signer except the right hand, which instead represents the woman's hand.

Because sign languages use multiple articulators and diagrammatic iconicity, different linguistic strategies (i.e., lexical signs, pointing, depicting constructions, CA) are not mutually exclusive and can be combined during encoding (Ferrara & Hodge, 2018; Perniss, 2007a). Note that subtle and reduced CA types, in Cormier, Smith, et al.'s (2015) terms, include the use of other linguistic strategies together with CA. For example, a signer can use CA to encode a referent with bodily articulators (e.g., eye gaze, facial expression, torso) and articulate a lexical sign on one or both hands to encode an action. Such combinations may be particularly useful for encoding transitive actions in relation to their patients, e.g., kissing the cheek of the child, considering that some lexical signs (so-called *directional verbs* or *indicating verbs*) can also make use of the body to establish a diagrammatic relation with components of CA (Cormier, Fenlon, et al., 2015). Thus, even in instances where different linguistic strategies are used for different articulators, the addressee has no problem decoding them because each piece of semantic information that is encoded by a specific articulator is decoded in a diagrammatic relation to all the other articulators employed. Or in other words, each articulator is embedded in a larger representation which constitutes a sum of meanings accessible through the articulators used and their relation to each other. The fact that multiple articulators can be linked together to simultaneously encode multiple semantic information units in a single construction provides a clear opportunity for communicative efficiency considering that related meanings can be encoded together to form a larger representation. Indeed, Slonimska et al., (2020) showed that signers exploit simultaneous and iconic constructions with the increasing information demands. It is therefore highly probable that the iconic properties of CA described above, including the possibility of the combination of CA with other strategies, are used for achieving efficient communication in sign languages.

3.1.2. Constructed action and informativeness

Until now, previous research has overwhelmingly concentrated on CA use in narratives (Cormier, Smith, et al., 2015; Hodge et al., 2019; Hodge & Johnston, 2014; Jantunen, 2017; Pizzuto et al., 2006, among others). The only two studies comparing CA use in narratives and other communicative contexts seem to indicate that in narratives CA occurs considerably more frequently. Sallandre et al. (2019) reported that the use of CA (called *transfer of person* or *double transfer* in their study) in LSF (French Sign Language) in narratives amounted to approximately 50% of all strategies used, while in a dialogue corpus it was only 7%, in an argumentative corpus it was 15%, and in recipe descriptions it was 27%. In line with the findings on the dialogue data, Ferrara (2012) found that in an Auslan (Australian Sign Language) conversation corpus CA was used six times less in comparison to narrative data, which led Ferrara to conclude that CA "should not be considered necessary, but that it is exploited in narrative contexts" (2012, p. 212). However, Quinto-Pozos (2007a, 2007b) found that in a movie clip description task, signers of ASL were likely to use CA and could not come up with other possibilities for encoding specific meaning when presented with stimuli of the animate entities involved in an action. Moreover, perceivers rated CA use as being clearer and more appropriate. Quinto-Pozos (2007a) argued that when encoding information about animate entities, CA "provides, in a simultaneous fashion, information that cannot be provided efficiently or robustly by using only signs or polycomponential signs" (p. 464) and that the prevalent iconicity and the possibility of one-to-one mapping between the body of the signer and the referent might prove to be a defining factor in the obligatory nature of CA in specific instances. To summarize, it appears that the need to use CA may vary depending on different contexts, and on the requirements that come with them, as well as on the type of the stimuli.

While narratives appear to be the most obvious context for eliciting CA, the fact that it has also been found outside narrative contexts, and that it even appears to be preferred over other linguistic strategies in some instances, might indicate that it is used to communicate information efficiently in its own right. Indeed, the referential value of CA has been acknowledged with regard to visibly depicting referents, indexing referents in space, and discourse cohesion (Cormier, Fenlon, et al., 2015; Cormier, Smith, et al., 2015; Liddell, 2003; Winston, 1991). For example, some research has shown that, while lexical signs are used to introduce referents in a story, CA is used more than lexical signs to maintain and/or reintroduce the referents (Cormier, Smith, & Sevcikova, 2013; Frederiksen & Mayberry, 2016; Hodge et al., 2019; Özyürek & Perniss, 2011; Perniss & Özyürek, 2015; Pizzuto et al., 2006). Recently, research on narrative data has shown that CA can function as the "sole conveyer" of information, i.e., encoding the core argument and predicate elements in a clause (Ferrara & Hodge, 2018; Ferrara & Johnston, 2014; Hodge & Johnston, 2014; Jantunen, 2017), leading some authors to suggest that "CA can function similarly to linguistic signs as a […] predicate and arguments" (Ferrara & Johnston, 2014, p. 204).

While Quinto-Pozos' (2007a, 2007b) research indicates that the referential properties of CA can also be taken advantage of for efficient communication outside narrative contexts, the design of that study did not allow this assumption to be assessed. In the same vein, while studies based on narrative corpora indicate that CA may indeed function as the carrier of the core information and not solely as an evaluative device, the narrative context might prove to be problematic for such an inquiry and conclusions. In the next section, we argue why the assessment of CA should go beyond narrative context in order to truly understand the referential capacity of this depicting strategy.

3.1.3. What narratives can and cannot tell us about the function of CA

Narratives require mastery of two functions: referential and evaluative (Labov & Waletsky, 1967). The referential function serves to make sense of the story and can be considered "a straightforward report of what occurred" (Cortazzi, 2014, p. 44). The evaluative function, on the other hand, serves to "[establish] some point of personal involvement" (Cortazzi, 2014, p. 44), which in turn implies the intensification of the factual information in the story with additional linguistic and paralinguistic strategies. For example, a signer might add a depiction to emphasize how a dog actually runs by using an excited facial expression with their tongue out, representing the emotional state of the dog during the action. Thus, the referential and evaluative functions are so intertwined in narratives that it becomes impossible to distinguish which linguistic strategy is used for which purpose. Curiously, some research has suggested that CA is used for evaluative function, i.e., to add color to the content, to make it more entertaining or vivid, and to capture the attention of the addressee (e.g., Dudis, 2002; Levelt, 1981; Mather & Winston, 1998; Poulin & Miller, 1995; Roy, 1989; Wilson, 1996; Winston, 1992). Although there are some recent studies that argue for the referential function of CA in narratives, they do not show whether it is used primarily for informative rather than evaluative function.

Note also that *addressee/recipient design* (i.e., adjusting the message by taking into account the needs of the addressee; Campisi & Özyürek, 2013; Clark, 1996) is radically different in narratives compared to purely informative tasks. In narratives, the goal is to tell a story and to be captivating and interesting while delivering information. Thus, the evaluative function is used deliberately. In cases where the goal of communication is efficient information transmission, the referential function is mainly required. Efficient communication can be interpreted according to Grice (1975) cooperative principle, where interlocutors have to be as informative as possible but also as concise as possible in transmitting information. As the amount of information that needs to be communicated increases, communicators are faced with the ever-growing challenge of accommodating the communicative needs of the addressee as well as their own. As a result, they are likely to adopt the most efficient strategy for doing so. Hypothetically, if the use of CA could be observed in such a setting, that is, when the information to be encoded increases in an experimentally controlled manner and communicators need to be efficient, it would be a strong indicator that it is not only used for referential purpose but also for achieving efficient communication.

3.2. The present study

In the present study we undertake to explore whether CA is used in a referential function in order to achieve efficient communication. We used a design (the same as in Slonimska et al., 2020) which reduces the confound of the evaluative function by presenting participants with a purely informative task of increasing demand with regard to the amount of information that has to be communicated. In such a task, the only requirement is to communicate the event's referential information. If we observe that signers increase their use of CA as a function of the increasing information load, we would have a strong argument for the referential use of CA. We hypothesize that signers will not opt to exclusively use lexical signs but instead will also use CA alone or in combination with other strategies in an informative task. Furthermore, we hypothesize that, as the amount of information that has to be encoded increases, so does the use of CA.

3.3. Methods

We used the video data collected by Slonimska et al. (2020). Here we report the design of the Slonimska et al. study in a shortened form and elaborate on the data coding scheme developed for the present study. The study was approved by the Ethics Council of the Institute of Cognitive Sciences and Technologies, CNR, Rome (protocol n. 0012633/2019).

3.3.1. Participants

Data was collected from 23 deaf adult participants (12 females, M-age = 30.5, range 18–57). Seventeen participants were native signers of LIS, all children of deaf parents.

Six participants were children of hearing parents and acquired LIS between ages 4-8 at school. All participants were daily users of LIS and reported it as their primary language for communication. Given some differences in regard to age of acquisition of the participants, we account for it in the analyses.

3.3.2. Material and design

The elicitation material for the experiment consisted of 30 unique images that represented an event involving two animate referents (there were six different referent pairs with 5 information density levels; e.g., referent pair: bird and bunny in Figure 3-2).

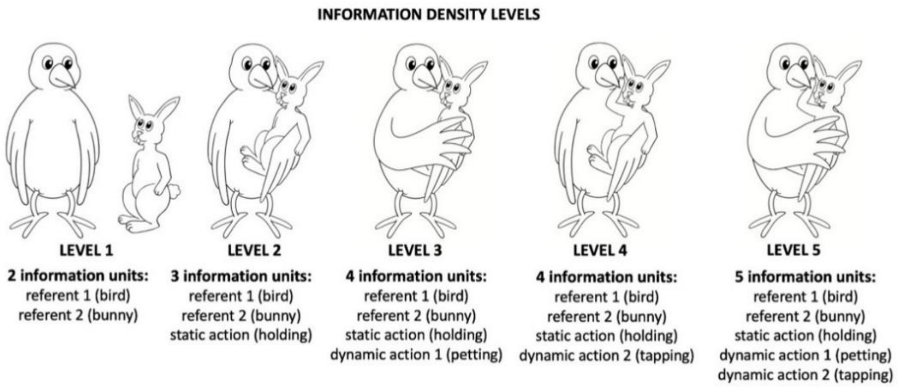


Figure 3-2. Stimuli of the images representing event of various semantic information density levels (referent pair: bird and bunny). Levels 1-2 are in JPG format, and levels 3-5 are in GIF format where only dynamic action is animated.

All stimuli represented animate referents in order to give signers the opportunity to opt for CA as an encoding strategy, considering that CA enables signers to give referents agency (Hodge et al., 2019). The images were divided across five levels (the images for levels 1–2 were in JPG format and the images for levels 3–5 were in GIF format), with each consecutive level representing an increase in the information density of the event. There were a minimum of two and maximum of five information units that needed to be encoded. Note that the number of information units in level 4 is the same as in level 3, but in level 4 both referents are simultaneously agent and patient, as opposed to the single agent and patient in level 3. Accordingly, level 4 increased in terms of perceptual complexity relative to level 3, but not in terms of information density. In all images both referents were represented as looking at each other, but because this information was not manipulated and remained constant across all levels, it was not considered in the encoding. Our aim was to use non-linguistic stimuli to elicit linguistic encoding in order to approximate as closely as possible situations in everyday life, in which people use language to describe events happening in the world. The format of the drawings (i.e., JPG or GIF) was chosen in order to sufficiently control the detail of each stimulus and assure that all stimuli were homogeneous.

3.3.3. Procedure

The participant was informed that they were about to play a *director–matcher* game in which they would play the role of *director* and another player, (a deaf confederate, native signer of LIS) who was seated facing the participant (see detailed set-up in Slonimska et al., 2020, p. 7), was assigned the role of *matcher*. The participant's task was to look at the images appearing on a screen one by one and in a semi-randomized order, and describe these images to the matcher, who would choose the correct image on a laptop. Before the experimental stimuli were presented, images with each referent were presented separately, one by one. The participant was invited to identify and describe these referents to the matcher. Once all the referents had been identified, the experimenter proceeded with the presentation of the experimental stimuli. The confederate always replied with positive feedback (e.g., signs for OK; yes; got it) after the images were described. Considering that participants were in a goal-oriented setting, they were expected to adopt a communicative strategy that was as efficient as possible when faced with increasing information demands in order to ensure that communication had been successful and their descriptions understood. After all the images had been described, the experimenter debriefed the participant about the experiment and answered the questions, if any were raised.

3.3.4. Coding

The video-recorded data was coded in the multimodal data annotation software ELAN, developed by Max Planck Institute for Psycholinguistics (Wittenburg et al., 2006). The duration of the videos was 7.23 min on average (SD = 1.39).

To determine the sequential organization of the production, we used the segmentation criteria set out in Slonimska et al. (2020, pp. 7–8): "Data segmentation was based on when a new movement of the signer's hand(s) started and ended, i.e., a stroke that could also be preceded by preparation, following Kendon (2004)." The start of a new segment was delimited by the start of the new movement of the other hand (i.e., preparation or stroke), and the new movement segment (MS) could also include the holding of the previous movement (see Figure 3-3, MS3–5). Coders annotated the presence of non-manual movements (change in torso position, head position, facial expression, eye gaze direction) in each movement segment. Therefore, a movement segment is determined by a change in at least one hand movement. Furthermore, if two hands were used to produce independent signs at the same time (e.g., CA with the left hand and pointing to self as the referent with the right hand; see Figure 3-3, MS2), that was coded as a single movement segment.

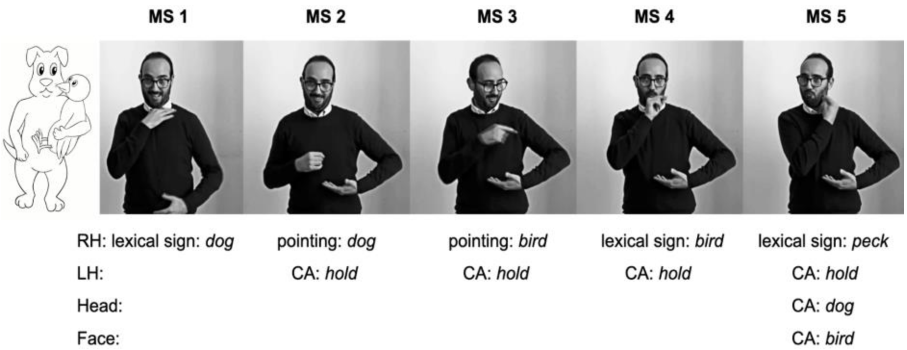


Figure 3-3. Example of the segmentation of a single stimulus with 5 MS and coding of linguistic strategy used in each MS (*Information density level 4*, referent pair: dog and bird).

All movement segments that were clear disfluencies or mistakes that signers corrected themselves were excluded from the analyses. Moreover, given that our focus was on how signers encoded the five information units that we manipulated in the different levels, we excluded additional movement segments that added extra information that was not the focus of our study (e.g., the size or shape of the referents, or movement segments encoding only the eye gaze direction of the referents). We then proceeded to assess how each movement segment was constructed with regard to the linguistic strategy used.

Each movement segment was coded with regard to the linguistic strategy or strategies it contained. First, we coded the general linguistic strategy of each movement segment: *lexical unit*, *constructed action, depicting construction, pointing,* and *combined* (for combined strategy, we noted which strategies were used in combination):

Lexical unit (LU) – a conventionalized sign with a fixed meaning, roughly comparable to words in spoken language;

Constructed action (CA) – a depicting strategy where the signer adopts a character's perspective of the event and maps a referent and its actions onto his own body;

Depicting Construction (DC) – a depicting strategy where the signer adopts an observer's perspective of the event, which is depicted in miniature scale in the signing space;

Pointing – use of index finger or palm for deixis;

Combined:

- *Lexical unit + Constructed action (LU + CA)*
- *Lexical unit + Depicting Construction (LU + DC)*
- *Lexical Unit + Pointing (LU + Point)*
- *Pointing + Constructed action (Point + CA)*
- *Pointing + Depicting construction (Point + DC)*
- *Constructed action + Depicting construction (CA + DC)*

Considering that it is often impossible to distinguish the handling and enactment of lexical signs (e.g., to pet; to hold) from CA (Cormier et al., 2012; Cormier, Smith, et al., 2015; Cormier, Smith, & Zwets, 2013; Ferrara & Halvorsen, 2017), we followed Cormier, Smith, et al. (2015) and coded the signs as *lexical units* if they were produced in an exclusively citational form as demonstrated on the *Spread the sign* webpage (www.spreadthesign.com), and/or they were available in the LIS–Italian dictionary by Radutzky (1992) and/or we were instructed by deaf informants.

We coded a MS as a CA if a referent and/or its actions were enacted. If two referents or the actions of different referents were enacted through CA, the linguistic strategy for encoding was also coded as *CA*. In order to determine whether non-manual articulators were used to encode the referents via CA, we followed the criteria for detecting CA in Cormier, Smith, et al. (2015). We coded for eye gaze if it was used for the purposes of enactment. If eye gaze to pointing or depicting signs was present during CA we did not code it as marker of CA and instead considered it referential eye gaze. With regard to the *combined* strategy, we noted combinations of linguistic strategies in each movement segment. For example, if CA was used simultaneously with a lexical unit, it was coded as *CA + LU*. If CA was used with pointing it was coded as *CA + Pointing*.

3.3.5. Reliability

All data was initially coded by the first author of the study. All coded data was controlled by a deaf researcher, a native signer of LIS. Another native signer of LIS coded 20% of the data. Reliability for linguistic strategy coded for each movement segment was very strong as revealed by Cohen's $\kappa = 0.90$.

3.3.6 Analyses

We analyzed the data in R using the lme4 package (Bates et al., 2015). We used the method of generalized mixed effect models (family=binomial) to test the effect of the *Information density level* on the *linguistic strategy* chosen for the encoding. The following random effects were considered for the baseline model: random intercept for *trial* (i.e., stimulus), random intercept for *participant*, random intercept for *referent pair*, and random slope for *Information density level by participant*. The following fixed effects were considered: *gender, age, age of LIS acquisition,* and *handedness*. The final baseline model was determined based on the best fit as revealed by ANOVA tests, or alternatively on the maximal random effects structure that converged in the model, following Barr et al. (2013). The best fit baseline model for each analysis is reported in the respective paragraph. Hierarchical contrasts between the levels were attained by re-levelling the primary model.

3.4. Results

We tested whether participants varied the proportion of specific linguistic strategies as a function of the increasing amount of information to be communicated. We hypothesized that participants would increase their use of CA and combine CA with other linguistic strategies (lexical units, depicting constructions, pointing) as the amount of information that had to be encoded increased. Given the type of the events represented in the stimuli (i.e., animate referents interacting with each other), we did not expect frequent use of depicting constructions.

Figure 3-4 presents the results with regard to the linguistic strategies participants used to encode stimuli for each information density level. Pointing and depicting constructions were scarcely used, and therefore in the analyses we concentrate on the three dominant strategies: lexical units, CA, and the combined strategy (simultaneous use of different strategies).

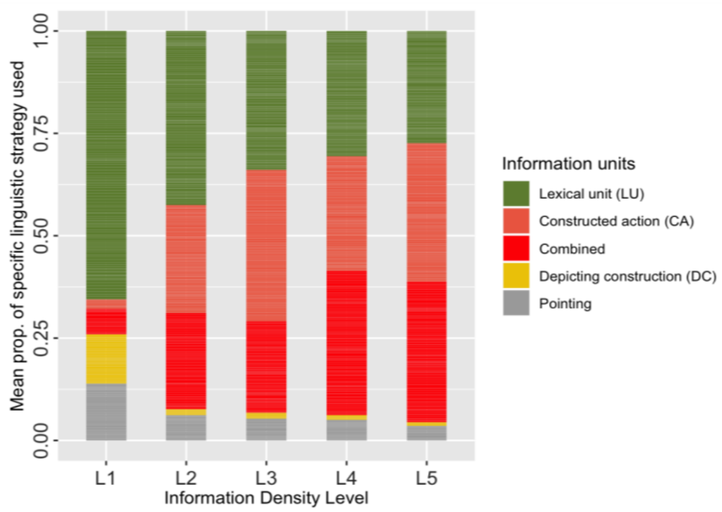


Figure 3-4. Raw proportions of linguistic strategies used to encode a stimulus in each information density level.

3.4.1. Lexical units

In the present analysis, we assessed how the use of lexical units was distributed across the information density levels. The baseline model consisted of the random effects of *participant* and *trial.* The outcome variable was the proportion of *lexical units* used: movement segments encoded via lexical units versus total number of coded movement segments used per encoding each stimulus. *Information density level* was compared to the baseline model, which consisted of random effects of *participant* and *trial,* and revealed a significant main effect ($\chi2(4) = 70.27$, $p<.001$). The primary model was releveled in order to attain hierarchical contrasts between the levels. Pairwise comparisons revealed that there was a significant gradual decrease in use of lexical units.

The strategy of using lexical units was used significantly more ($\beta=-0.82$, SE=0.13, CI [-1.07; -0.57], $z=-6.37$, $p<.001$, see Table 3-1 for a summary of the model) in level 1 (M=0.66, SD=0.19) than in level 2 (M=0.43, SD=0.16). In level 2 it was used significantly more ($\beta=-0.38$, SE=0.12, CI[-0.61; -0.15], $z=-3.22$, $p=.001$) than in level 3 (M=0.34, SD=0.12), while in level 3 it was comparable ($\beta=-0.13$, SE=0.11, CI[-0.58; 0.10], $z=-1.12$, $p=.262$) with level 4 (M=0.31, SD=0.10), and level 4 was comparable ($\beta=-0.18$, SE=0.11, CI[-0.393; 0.032], $z=-1.67$, $p=.096$) with level 5 (M=0.27, SD=0.11).

Random effects	**Variance**		**SD**			
Trial	0.00		0.00			
Participant	0.196		0.443			
Number of obs: 678	Groups: Trial=30,		Participant=23			
Fixed effects	**95% CI**		**B**	**SE**	***z***	***p***
	Lower b.	Upper b.				
(Intercept)	0.18	0.70	0.44	0.13	3.34	**.001**
Level2	-1.07	-0.57	-0.82	0.13	-6.37	**<.001**
Level3	-1.44	-0.96	-1.20	0.12	-9.72	**<.001**
Level4	-1.57	-1.08	-1.33	0.12	-10.71	**<.001**
Level5	-1.74	-1.27	-1.51	0.12	-12.59	**<.001**

Table 3-1. Best fit model in a logit scale (model fit by maximum likelihood, Laplace Approximation) regarding the proportion of Lexical units used for encoding. Contrasts reflect pairwise comparisons between level 1 and all other levels.

3.4.2. Constructed action

Next, we explored the effect of information density level on the use of CA (i.e., proportion of movement segments encoded via CA versus total number of coded movement segments used per encoding each stimulus). We compared the fixed effect of *Information density level* to the baseline model, which consisted of random effects of *participant* and *trial* (see Table 3-2 for a summary of the model). There was a significant effect of *Information density level* ($\chi2(4)$ =68.68, p<.001). CA was used significantly more (β=7.93, SE=-0.31, CI[1.83; 3.03], z=2.32, p<.001) in level 2 (M=0.26, SD=0.7) than in level 1 (M=0.02, SD=0.05.) and significantly more (β=0.48, SE=0.18, CI[0.14; 0.82], z=2.73, p<.006) in level 3 (M=0.37, SD=0.08) than in level 2. However, in level 3 CA was used significantly more (β=-0.43, SE=-0.17, CI [-0.43; -0.75], z=-2.53, p<.011) than in level 4 (M=0.28, SD=0.08). Levels 4 and 5 (M=0.34, SD=0.07) were comparable (β=0.28, SE=-0.16, CI [-0.04; 0.60], z=1.71, p<.087).

Random effects	**Variance**		**SD**			
Trial	0.047		0.220			
Participant	0.060		0.244			
Number of obs: 678	Groups: Trial=30,		Participant=23			
Fixed effects	**95% CI**		**B**	**SE**	***z***	***p***
	Lower b.	Upper b.				
(Intercept)	-4.04	-2.93	-3.49	0.28	-12.37	**<.001**
Level2	1.83	3.03	2.32	-.31	7.93	**<.001**
Level3	2.32	3.50	2.91	0.30	9.67	**<.001**
Level4	1.89	3.08	2.48	0.30	8.23	**<.001**
Level5	2.18	3.35	2.76	0.30	9.26	**<.001**

Table 3-2. Best fit model in a logit scale (model fit by maximum likelihood, Laplace Approximation) regarding the proportion of CA used for encoding. Contrasts reflect pairwise comparisons between level 1 and all other levels.

The results confirmed our hypotheses that signers are likely to increase use of CA as the amount of information that has to be encoded increases. For example, in level 2 signers could encode both referent 1 and its static action by means of CA (Figure 3-5), and in level 3 they could encode referent 1, its static action and add a dynamic action in a single CA (Figure 3-6). In levels 4 and 5 they could encode referent 2 and its dynamic action 2 (Figure 3-7) or, alternatively, referent 1, its static action and referent 2's dynamic action via CA (Figure 3-8). Encoding both referents and all three actions was not attested in our data.

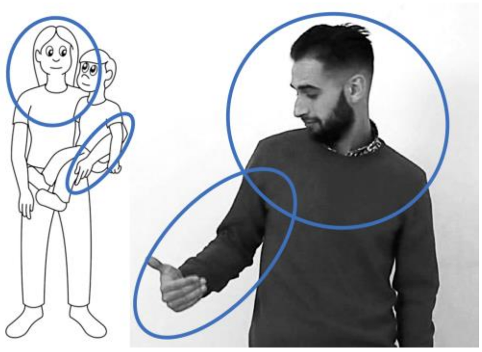


Figure 3-5. A signer depicting referent 1 – *woman* (encoded through head direction, facial expression, and eye gaze) and the static action (the signer's right hand) via CA (*Information density level 2*).

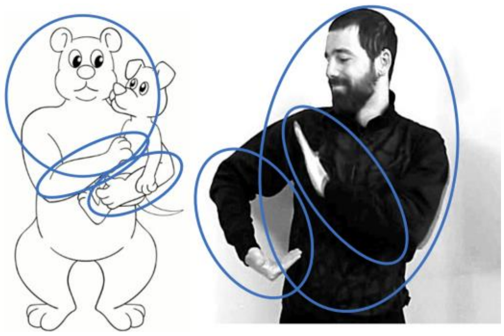


Figure 3-6. A signer depicting referent 1 - *bear* (encoded through torso, head, eye gaze, and facial expression), his static action (the signer's right hand) and dynamic action 1 - *petting* (the signer's left hand) via CA (*Information density level 3*).

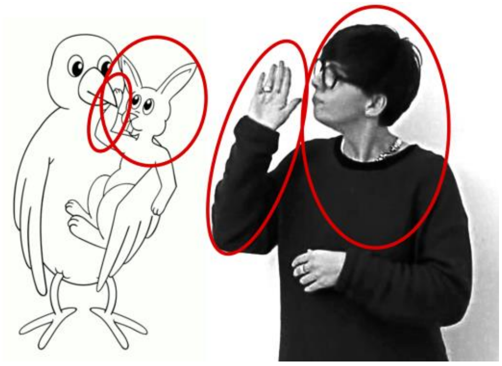


***Figure 3-7**.* A signer depicting referent 2 – *bunny* (encoded through torso, head, eye gaze, and facial expression) and dynamic action 2 – *tapping* (the signer's right hand) via CA (*Information density level 4*).

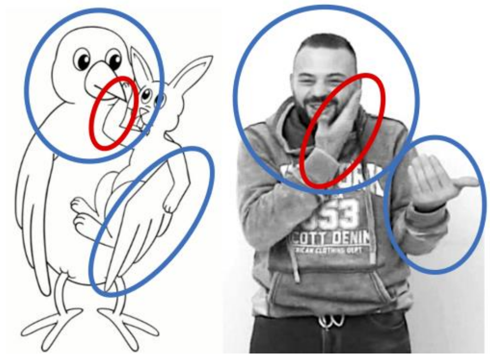


Figure 3-8. A signer depicting referent 1 – *bird* (encoded through torso, head, facial expression) and its static action (the signer's left hand) and dynamic action of referent 2 (the signer's right hand) via CA (*Information density level 4).*

We note that we found a discrepancy with regard to level 3, where CA was used significantly more than in level 4, contrary to what was expected. In level 3, all the information about referent 1, the agent of the actions, can be given in a single CA in which the signer's entire body represents referent 1 and its actions can be easily mapped through CA onto the signer's body parts (Figure 3-6). In level 4, referent 1 is the patient of the action produced by referent 2 (Fig.3-7, 3-8). Inspection of the data showed that the action performed by referent 2 could be encoded not only by means of CA but also LU or a depicting construction. We explore the *combined* strategy in the next section.

3.4.3. Combined strategies

We assessed how combinations of different linguistic strategies (i.e., proportion of movement segments encoded via a combination of multiple linguistic strategies versus total number of coded movement segments used per encoding each stimulus) were distributed across the information density levels. We compared the fixed effect of *Information density level* to the baseline model, which consisted of a random effect of *participant* and *trial* (see Table 3-3 for a summary of the model).

Random effects	**Variance**		**SD**			
Trial	0.13		0.35			
Participant	0.49		0.70			
Number of obs: 678	Groups: Trial=30,		Participant=23			
Fixed effects	**95% CI**		**B**	**SE**	***z***	***p***
	Lower b.	Upper b.				
(Intercept)	-3.44	-2.36	-2.90	0.28	-10.51	**<.001**
Level2	1.07	2.22	1.64	0.29	5.64	**<.001**
Level3	1.05	2.17	1.61	0.29	5.60	**<.001**
Level4	1.63	2.75	2.19	0.29	7.68	**<.001**
Level5	1.67	2.77	2.22	0.28	7.85	**<.001**

Table 3-3. Best fit model in a logit scale (model fit by maximum likelihood, Laplace Approximation) regarding the proportion of combined strategies used for encoding. Contrasts reflect pairwise comparisons between level 1 and all other levels.

There was a significant effect of density level ($\chi2(4)$ =44.041, $p<.001$). Combined strategies were used significantly more (β=1.64, SE=0.29, CI [1.073; 2.22], z=5.64, $p<.001$) in level 2 (M=0.23, SD=0.19) than in level 1(M=0.06, SD=0.09). Level 2 and level 3 (M=0.22, SD=0.15) were comparable (β=-0.03, SE=0.24, CI [0.51;0.44], z=-0.14, p=.89). Combined strategies were used significantly more (β=0.58, SE=0.24, CI [0.12; 1.04], z=-0.25, $p<.01$) in level 4 (M=0.35, SD=0.12) than in level 3, though use in level 4 (β=0.03, SE=0.23, CI [-0.42; 0.48], z =0.13, p =.09) was comparable to level 5 (M=0.34, SD=0.13).

When we explored the use of each type of combination, we found that CA was combined with another strategy (LU, DC, pointing) almost exclusively (96%) except in level 1, where CA + another strategy constituted 65% (Figure 3-9) of combined strategies.

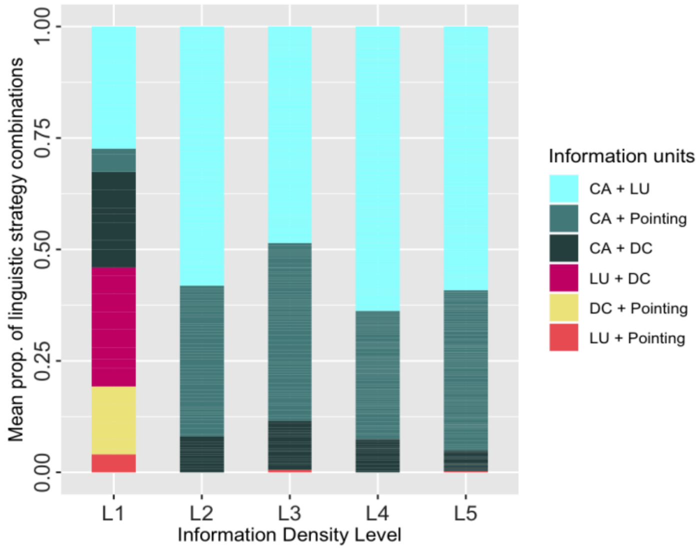


Figure 3-9. Raw proportions of the linguistic strategy combinations used for encoding a stimulus in each level.

In level 2 and level 4 only CA + another strategy was used, while in level 3 and level 5 only one instance of a combination that did not contain CA occurred. Accordingly, the *combined* strategy was predominantly used to combine CA with another linguistic strategy for encoding; in the majority of cases, CA was combined with lexical signs, followed by combination with pointing in levels beyond 1.

As mentioned above, CA as a single strategy was used significantly more in level 3 (referent 1, referent 2, static action, dynamic action 1) than level 4 (referent 1, referent 2, static action, dynamic action 2). However, we also found that there was a significant increase in the combined strategy between levels 3 and 4. In other words, in level 4 signers used more CA in combination with another linguistic strategy. Data examination revealed that regardless of the fact that signers could use full CA in cases where action was produced by the mouth and head articulators (as in *licking, kissing, pecking*) by mapping the articulators of the referent onto the signer's articulators, they nevertheless chose to encode it through the hand by means of lexical sign or depicting construction. Such combined strategies consisted of encoding referent 2 via CA and its action via another strategy, e.g., depicting construction (Figure 3-10).

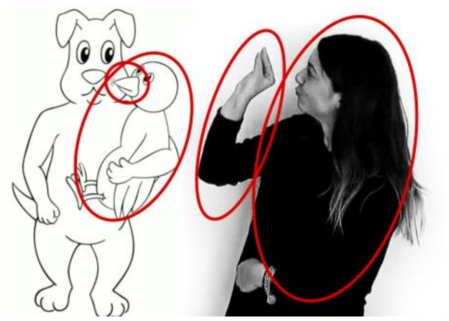


Figure 3-10. A signer encoding referent 2 –*bird* (encoded through the torso, head, eye gaze, and facial expression) and dynamic action 2 –*pecking* (the signer's right hand) with depicting construction (*Information density level 4*).

Alternatively, it could be used to encode the actions of both referents: while CA was used to encode referent 1 and the holding action, one of the hands was partitioned off in order to encode the action of the referent 2 via lexical sign (Figure 3-11). Interestingly, some signers would accompany the action sign encoded by the hand with non-manual articulators as well, but they never used non-manual articulators only. For example, when encoding a bird mapped on the body through a torso shift and the pecking action with a depicting construction or lexical sign, a signer would also map the beak of the bird by pursing her lips and moving her head back and forth to reproduce the pecking action (Figure 3-10). Some signers, however, did not do this, indicating that redundancy in action encoding is to some extent a feature of a signer's individual style.

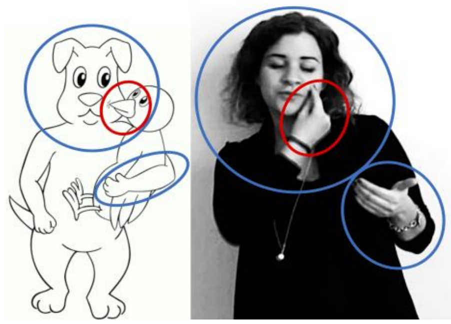


Figure 3-11. A signer encoding referent 1 –*dog* (encoded through the torso, head, and eye gaze) and holding action (the signer's left hand) via CA and dynamic action 2 –*pecking* (the signer's right hand) with a lexical sign (*Information density level 4).*

3.5. Discussion

In the present study we hypothesized that if CA in sign languages can serve a primarily referential function, we would see an increase in their use when the main goal of the task was efficient communication. Assessing CA in a controlled experimental setting allowed us to reduce the need for signers to use the evaluative function and instead focus primarily on referential function. Furthermore, the design of the study allowed us to assess not only whether CA was used for referential purposes, but whether its properties of implementing diagrammatic iconicity to encode multiple event elements simultaneously was also used to achieve communicative efficiency when faced with increasing information demands. Our results revealed that CA (also in combination with other strategies) was the prevalent strategy used in all levels except in level 1. We also found that, as the amount of information that needed to be encoded increased, CA (alone and in combination with other strategies) either increased or was comparable to the preceding level (e.g., levels 4 and 5). An exception to this finding was more use of CA alone in level 3 than in level 4. Yet the use of CA combined with other strategies was found more in level 4 than in level 3. We address these findings below.

3.5.1. CA as a referential device

In our data, we found an overwhelming amount of CA use, both as an independent strategy and in combination with another linguistic strategy, such as lexical signs, pointing, and depicting constructions. Thus, the present findings corroborate previous research arguing that CA is an integral part of sign languages and that it can be used for referential purposes. The general tendency to increase the use of CA (apart from levels 4 and 5) as well as the use of CA in combination with another strategy indicates that this strategy and its use in combination with other strategies can be employed to achieve communicative efficiency. Note that the design of the study was based on the increase of one information unit in each consecutive level (except levels 3 and 4, which differed in perceptual complexity and not information density), which might have been too small of a difference to detect the significant effect between all levels we compared. In addition, it is also possible that different information units (e.g., different types of action in our study) have their own constraints on whether CA can or cannot be used for referential purposes. Indeed, exactly this factor appears to explain the unexpected finding of CA alone being used less in level 4 compared to level 3. We address this finding later on in this Discussion.

Given that even languages that have another primary communication channel (i.e., voice) resort to iconic gestures in some instances to communicate efficiently (e.g., Campisi & Özyürek, 2013; Holler & Wilkin, 2011), it appears only logical to assume that, in languages that employ a visual channel exclusively, depiction would also play a crucial role in information transmission – a view that has been rapidly gaining prominence in sign language research (Cormier, Smith, et al., 2015; Ferrara & Hodge, 2018; Hodge & Johnston, 2014; Jantunen, 2017; Puupponen, 2019). Indeed, the act of combining depictive properties with more discrete conventionalized properties for linguistic purposes appears to be a rather sophisticated task as exemplified by research demonstrating that the use of CA together with other strategies is particularly hard for children to acquire (BSL: Cormier, Smith, & Sevcikova, 2013; LIS: Slonimska et al., 2018). Assessment of how CA is used in combination with other strategies in various contexts can thus further our understanding of the interplay between the linguistic strategies that signers have at their disposal. An undertaking for future research would be to implement the same task design presented in this study to assess how increasing information demands influences the use of CA (alone and in combination) in children.

It is important to highlight that the consecutive and simultaneous interplay between highly conventionalized (i.e., lexical) and gradient iconic signs (i.e., CA) allows great flexibility in encoding. Specifically, the lexical signs for referents frame the use of CA so that it can be interpreted unambiguously (BSL: Cormier, Smith, & Zwets, 2013; Auslan: Hodge & Ferrara, 2014). Accordingly, once the referents are introduced via lexical signs, the signer can take advantage of the depictive properties of the language in order to encode the event more efficiently than would be possible if strictly consecutive encoding of one sign - one meaning were used.

The advantage of encoding multiple information units in a single construction has been shown in our recent study that used the same video data (Slonimska et al., 2020). Results revealed that, as the amount of information that had to be encoded increased, signers increased the simultaneous encoding of multiple units of information. These findings indicate that signers take advantage of the affordances of sign language to exploit its referential capacity to the fullest. The present study contributes to the findings by Slonimska et al. (2020) by illuminating linguistic strategies used to achieve communicative efficiency, something that has not been investigated so far. In our data, we found that a majority of the combined strategies included CA. This indicates that signers can use CA flexibly with different linguistic strategies to encode multiple information units in a construction. Considering that the body of the signer serves as a central coordinate point, each element can be interpreted in relation to each of the others (i.e., diagrammatic iconicity), forming a single but also complex representation in which multiple elements of the event can be depicted simultaneously (Slonimska et al., 2020). For example, in order to encode both referents, signers could use body partitioning (Dudis, 2004) instead of encoding information about each referent separately. Note that we observed that actions of the referent 2 that were not performed with the hand but that could be encoded via CA by the use of head or mouth articulators were nevertheless encoded with the signer's hand and thus via another strategy (either LU or DC). Exactly this observation could explain why CA alone was used less in level 4 compared to level 3. Namely, in level 4 a subset of the actions by referent 2 were not encoded by CA only but by CA in combination with another strategy, thus diminishing the overall proportion of CA as a sole strategy used in level 4 (see Section 3.4.3.). There might be some very practical reasons for signers using their hands to encode actions. If an action is encoded by the signer's hand it can then be added to the diagram in a meaningful way. That is, it can establish the relation between the action, its agent, and its patient. For example, to encode that referent 2 is pecking referent 1 on the cheek, the signer can use a manual sign for pecking by simply directing it to the body part where the action occurs (i.e., the cheek of referent 1). In contrast, the signer cannot encode referent 1 by mapping it onto their own body (i.e., torso and head) and at the same time encode referent 2's pecking action by using their mouth to establish the relation between referent 1 and referent 2. In other words, a signer directing their mouth to their own cheek is simply impossible from an articulatory viewpoint (i.e., *modal affordances*; Puupponen, 2019). The problem is solved, however, if the action performed by the mouth is encoded by the hand, which can then be easily directed to any part of the patient's body. In such an instance, body partitioning is useful for encoding information precisely and at the same time efficiently by explicitly keeping the patient of the action present (Dudis, 2004). Thus, in order to communicate efficiently, signers do not adhere to description by means of lexical signs only but can additionally take advantage of the rich resources of depictive properties which are in the repertoire of their language.

3.5.2. The quest of iconicity towards language: the lexical/discrete vs. gestural/gradient dichotomy

Although research has gone a long way in acknowledging the crucial role of iconicity for sign language organization, production, and processing (Perniss et al., 2010; Vigliocco et al., 2005, 2014), the leading view has nevertheless stressed the necessity of distinguishing between linguistic and gestural features (Duncan, 2005; Goldin-Meadow & Brentari, 2017; Quinto-Pozos & Mehta, 2010) or alternatively between discrete + conventional and gradient + unconventional features in sign languages (Cormier et al., 2012; Johnston & Schembri, 1999; Liddell, 2003). It is only relatively recently that researchers have started arguing for revisiting the juxtaposition of *arbitrary/categorical = lexical* versus *iconic/gradient = gestural/non-lexical* properties as a decisive factor of linguistic status (Clark, 2016; Dotter, 2018; Kendon, 2004, 2014; Müller, 2018). Furthermore, our findings suggest that the sharp distinction between depicting and describing functions of language should possibly be reconsidered (Clark, 2016; Clark & Gerrig, 1990).

The fact that the signers in the present study preferred using CA over lexical signs in an informative task indicates that some concepts that are rooted in human experience, like actions that are performed on a daily basis, e.g., holding different objects, do not require a lexicalization process in order to be included in the linguistic structure of the populations that rely heavily on such constructions, like signers. Indeed, Fuks (2014, p. 152) notes: "Arbitrariness in the visual gestural modality, by contrast, is a constraint resulting from the entity's features. That is, it is used in signed languages only in those cases when iconicity cannot be used […]." Hockett (1978), in comparing spoken and sign languages, notes that, while in spoken languages the majority of iconicity has to be squeezed out due to its linear organization (when speech only is considered), sign languages have the chance to maintain iconicity to a far greater extent. As a consequence, languages in both modalities adapt to work with what they have. In this respect, iconicity can be taken full advantage of not only for the imagistic iconicity of lexical signs that resemble their referents but also for both imagistic and diagrammatic iconicity when it comes to communicating about what a body looks like, what kind of actions it makes, and where it stands in relation to the world and phenomena around it. And given that bodily actions are entrenched in human experience, they do not pose a decoding problem, all the more so when embedded in context.

When coding the data, we were faced with ever-growing doubts about the correctness and feasibility of distinguishing between lexical signs and CA. Specifically, determining the cut-off point between lexical and depicting signs (such as in CA or DC) was sometimes quite problematic. For example, an action like petting turned out to be particularly difficult to classify, given that the handshape is similar to the actual action of petting (Figure 3-6). Using citational forms as a benchmark is not an optimal solution, given that the citational forms of verbs found in dictionaries serve more as an umbrella term for a concept (e.g., *take*) that, when used in context, is most often encoded via a specific verb in form of depiction specifying the referent, e.g., *take a book, take a pen* (Tomasuolo et al., 2020). The fact that categorizing signs is difficult even in an experimental setting with a highly controlled design goes against the view of bounded lexicality in sign languages and instead supports the concept of gradience. Accordingly, language should be viewed as a set of complex structures that speakers and signers can bring into play as appropriate and tweak, squeeze, and stretch according to need in order to transmit their desired meaning as truthfully as possible. In that respect, the possibility of gradience is precisely what allows language to be as rich as it is. Rather than dividing the world into two based on linguistic vs. gestural or discrete vs. gradient dichotomies, it would perhaps be wiser to consider it as a continuum with signs having the possibility of being used on a spectrum between the two categories, which are equally linguistic on both ends and in between (Ferrara & Halvorsen, 2017; Ferrara & Hodge, 2018; Hodge & Ferrara, 2014; Jantunen, 2017).

3.6. Conclusion

In the present study we found that CA was frequently used in an informative task, extending findings on the referential function of CA in a primarily informative and non-narrative context. Furthermore, we found that signers tended to use more CA and more often combined CA with other linguistic strategies as the amount of information that had to be encoded increased, that is to be communicatively efficient. Thus, we showed that depictions like CA can be used with referential function, which is usually considered to be achieved with descriptions. We argue that signers use CA for descriptive purposes, due to the efficiency afforded by imagistic and diagrammatic iconicity, which allows for the meaningful combination of multiple information units into a single representation. Language does not consist merely of words or signs that people organize in strictly linear structures, but rather it also consists of fascinatingly rich depictive strategies that can combine different levels of linguistic representation to transmit meaning.

3.7. Acknowledgments

We wish to thank the participants for their participation and authorization to publish their images. We also wish to thank Barbara Pennacchi for drawing the stimuli for the experiment and assistance with data collection. We wish to thank Alessio Di Renzo, Luca Lamano, Tommaso Lucioli, Alessandra Ricci, and Morgana Proietti for assistance with data collection, coding and visualizations. Furthermore, we wish to thank the Editor and three anonymous reviewers who have contributed to the paper with their valuable feedback and suggestions during the revision process.

Chapter 4

Simultaneity as an emergent property of efficient communication in language: A comparison of silent gesture and sign language

This chapter is based on:

Slonimska, A., Özyürek, A., & Capirci, O. (2022). Simultaneity as an emergent property of efficient communication in language: A comparison of silent gesture and sign language. *Cognitive Science, 46*(5), e13133.

Abstract

Sign languages use multiple articulators and iconicity in the visual modality which allow linguistic units to be organized not only linearly but also simultaneously. Recent research has shown that users of an established sign language such as LIS (Italian Sign Language) use simultaneous and iconic constructions as a modality-specific resource to achieve communicative efficiency when they are required to encode informationally rich events (Slonimska et al., 2020). However, it remains to be explored whether the use of such simultaneous and iconic constructions recruited for communicative efficiency can be employed even without a linguistic system (i.e., in silent gesture) or whether they are specific to linguistic patterning (i.e., in LIS). In the present study, we conducted the same experiment as in Slonimska et al. (2020) with 23 Italian speakers using silent gesture and compared the results of the two studies. The findings showed that while simultaneity was afforded by the visual modality to some extent, its use in silent gesture was nevertheless less frequent and qualitatively different than when used within a linguistic system. Thus, the use of simultaneous and iconic constructions for communicative efficiency constitutes an emergent property of sign languages. The present study highlights the importance of studying modality-specific resources and their use for linguistic expression in order to promote a more thorough understanding of the language faculty and its modality-specific adaptive capabilities.

4.1. Introduction

One of the defining properties of natural languages is segmenting holistic representations into smaller meaning units that can be combined into larger meaning units, allowing compositionality (Goldin-Meadow et al., 1996; Kirby et al., 2008). This has been shown to constitute an emergent property of linguistic systems to accommodate the pressures of communicative efficiency during language use and language transmission to new learners (Kirby et al., 2008; Motamedi et al., 2019; Senghas et al., 2004). One of the pieces of evidence for this claim comes from sign language emergence research which has shown the emergence of segmentation out of initially holistic forms and linear sequencing of these segmented meaning units (Senghas et al., 2004). However, in sign languages, due to the affordances of the visual modality to use multiple articulators (i.e., hands, torso, head, facial expression, eye gaze) and iconicity, i.e., "the existence of a structure-preserving mapping between mental models of linguistic form and meaning" (Taub, 2001, p. 23), meaning units can be organized not only linearly but also simultaneously. Namely, multiple articulators can be used to encode different semantic information units simultaneously and diagrammatic iconicity in particular can be used to establish a motivated relationship between these simultaneously encoded units (Perniss, 2007; Risler, 2007; Slonimska et al., 2020). In the present study, we investigate if simultaneity, in addition to linearity, constitutes an emergent property of sign languages by comparing the use of simultaneous constructions in LIS (Italian Sign Language) to that of silent gestures used by hearing Italian speakers.

In a previous study, we showed that signers of LIS use simultaneous constructions to achieve communicative efficiency when they are asked to encode informatively rich events in a controlled interactive task (Slonimska et al., 2020). Namely, we found that signers increased their use of multiple articulators to simultaneously encode multiple units of information as the demands for communicative efficiency, that is the amount of semantic information units that needed to be communicated increased (Slonimska et al., 2020). We argued that signers can link simultaneously employed articulators into a coherent diagrammatically iconic representation to encode related units of meaning as closely to each other as possible (e.g., see Figure 4-1a below), in order to strive for *dependency distance minimization*. Dependency distance minimization is argued to lead to faster representation access and is used to achieve communicative efficiency in spoken languages (Gibson et al., 2019; Hawkins, 2004). We showed that simultaneity can serve as one of the modality-specific properties of clustering related meanings as close together as possible and thus achieving communicative efficiency in sign languages.

Yet, whether simultaneity constitutes an emergent linguistic property of sign languages that has evolved for communicative efficiency or whether it reflects a general expressive ability in the visual modality even outside of a linguistic system is not known. In the present study we aim to fill this gap. To do so, we follow up on our recent work (Slonimska et al., 2020) and compare the use of simultaneity in LIS to its use in silent gesture. Silent gesture is an experimental paradigm in which hearing adult participants are asked to use only their body and no speech to communicate and thus are required to use only their gestures to represent certain content. This research has found robust evidence that silent gesture is not dependent on the spoken language used by the participants but rather reflects representations based on shared visual-motor imagery (Gibson et al., 2013; Goldin-Meadow et al., 1996, 2008; Ortega & Özyürek, 2020; Özçalışkan et al., 2016; Schouwstra & de Swart, 2014). For example, analyses of silent gestures reveal that semantic elements constituting event components are ordered in similar linear structures across speakers of different languages (e.g., Gibson et al., 2013; Goldin-Meadow et al., 2008; Özçalışkan et al., 2016) and types of iconic representations are distinguished based on semantic category (e.g., Ortega & Özyürek, 2020). Yet, it is not known how people recruit simultaneity in silent gesture to represent complex events and whether and how it differs from simultaneous constructions used by signers.

A comparison of the use of simultaneity in silent gesture and LIS can result in two scenarios. One possibility is that when faced with increasing information demands both gesturers and signers will increase use of simultaneous constructions when they need to be more communicatively efficient. However, LIS signers will use more and/or qualitatively different simultaneous constructions than silent gesturers because the linguistic tools for doing so are built into the LIS system. This would show that simultaneity, despite being available as an affordance of communication in the visual modality for both groups, constitutes an emergent property in a linguistic system adapted for accommodating the pressure for communicative efficiency. An alternative possibility is that simultaneity can be recruited in silent gesture due to the natural affordances of the visual modality and be used to achieve communicative efficiency as often and/or even more than in LIS. This would indicate that simultaneity constitutes a general resource available also outside the linguistic system and as such can be recruited to attain communicative efficiency by signers and gesturers alike.

4.1.1. Language emergence and structural organization in the visual modality

Segmentation has been considered as one of the design features of language allowing emergence of compositional structure considering that "segmenting out one component of a simultaneous event allows the language user to combine that component with other elements, thus leading to new combinatorial possibilities not imaginable with the conflated [i.e., holistic] form alone" (Özyürek et al., 2015, p. 86). Experimental and computer simulation studies indeed show that communicative signal is likely to evolve from a holistic form in which "a signal stands for the meaning as a whole, with no subpart of the signal conveying any part of the meaning in and of itself" (Smith et al., 2003, p. 372) into segmented and compositional structure through conventionalization of the segmented meaning elements and their systematic recombination to create new meanings (Beckner et al., 2017; Kirby, 2000; Kirby et al., 2008, 2014, 2015; Motamedi et al., 2019; Nölle et al., 2018; Raviv et al., 2019; Theisen et al., 2010). These studies show that this process emerges as an adaptation to pressures of language use and transmission to new learners, which push languages towards becoming more communicatively efficient and easier to learn.

The same emergent trajectory has been observed in research on emerging sign languages in cases such as NSL (Nicaraguan Sign Language) and homesign systems developed by deaf children growing up without exposure to any conventional sign language input (Goldin-Meadow, 2015; Özyürek et al., 2015; Senghas et al., 2004, 2010, 2013). For example, studies on NSL have shown how the maturation of the linguistic system can be detected in a gradual increase of segmentability in motion events, which are by nature holistic with regard to expressing simultaneously occurring manner and path components (e.g., a cat rolls down a hill) (Senghas, 2019; Senghas et al., 2004, 2010, 2013). Senghas et al. (2004) found that within three cohorts of NSL users, each new cohort used significantly more segmented forms arranged linearly (i.e., separate signs for manner and path) for the motion events than the previous cohort. Furthermore, partially segmented forms (e.g., using two signs in which one sign conflated both manner and path while the other sign encoded only one of the elements), were found to be prevalent only in the first cohort of NSL signers while the second and third cohorts preferred fully segmented and linearized structure (Senghas, 2019; Senghas et al., 2004, 2010, 2013).

A study by Özyürek et al. (2015) on Turkish homesigners (i.e., deaf children with no access to a conventional language) further showed that even though homesigners did use partially segmented forms in describing motion events with manner and path, the prevalent strategy was to use conflated forms (i.e., holistic representation), that is, one gesture for both manner and path. Importantly, Özyürek et al. (2015) compared these data to productions of hearing adults in a silent gesture condition and found that silent gesturers used even more conflated forms than homesigners, indicating that while homesigners are not exploiting segmentation to the same extent as later cohorts in an emerging sign language, they have nevertheless embarked on the road to segmentation.

The above described studies have focused on the emergence of segmented linear structures allowing compositionality as a core emergent property of linguistic systems. However, a crucial factor to consider is that organization in the visual modality allows segmented forms to be organized not only linearly but also simultaneously (i.e., *simultaneous compositionality*). For example, Senghas & Littman (2004), who compared three cohorts of NSL signers to Nicaraguan and Spanish speakers and LSE (Spanish Sign Language) signers, showed that a likely trajectory of an emerging sign language starts with holistic expressions that gradually become more segmented to then be used in linear as well as simultaneous constructions. Their data revealed that while the second and third cohort signers expressed manner and path mainly in a linear sequence (as found in Senghas et al., 2004), signers of the established sign language, LSE, used simultaneous (i.e., manner and path in a single sign) rather than linear constructions. This use of simultaneous expressions was interpreted as a way to bring already segmented meaning elements together through simultaneous compositionality. Thus, while in the initial stages of language emergence simultaneity appears to be a feature of holistic representation in which manner and path are conflated in a single gesture, as language evolves it decomposes such holistic representations that conventionalize into separate linguistic units that can then be either combined linearly or simultaneously. Thus, while the final forms of such holistic expressions on the one hand and simultaneously compositional on the other can appear similar to each other, they differ in the way they have been constructed, i.e., holistically by silent gesturers and compositionally by signers. It is possible to speculate that over a few more generations, emerging sign languages like NSL might become more similar to established sign languages with respect to the encoding of motion verbs and converge on simultaneous representations of segmented elements of manner and path. Thus, simultaneity may be an emergent property of language.

This trajectory of moving from linear to simultaneous compositionality is also attested in developmental research on sign language acquisition in respect to the encoding of motion verbs (Meier, 1987; Morgan et al., 2002; Newport, 1981, 1988; Supalla, 1982). For example, studies on acquisition of ASL (American Sign Language) show that children acquire the morphological components of motion verbs in a piece-by-piece fashion, just like children of spoken languages do (Newport, 1981, 1988; Supalla, 1982). In the initial stages of acquisition, signing children appear to use only one element (e.g., only the path of the motion) from the complex adult-like form in which the path, the manner and the handshape of the referent are specified. In later stages, children stop omitting other elements of these complex verbs, but unlike adults, who encode these elements simultaneously, children encode them in a linear manner. It is not until about the age of 4–5 that they start producing adult-like simultaneous forms. A similar developmental trajectory has been also observed for complex verbs, where information is encoded linearly at the initial stages and gradually moves toward adult-like simultaneous forms (Meier, 1987; Morgan et al., 2002). Taken together, developmental research on sign language acquisition indicates time is needed to master simultaneous constructions. This suggests that such forms might be emergent properties in linguistic systems as well.

The studies mentioned above explored the structural possibilities for encoding motion event components either linearly or simultaneously by means of a single articulator (e.g., one hand representing the manner of rolling as it moves downwards to represent path). However, another way sign languages can represent event components simultaneously is through the use of multiple articulators. Signers can exploit both hands, their torso, head, facial expression, and eye gaze when encoding multiple semantic elements of an event in order to represent simultaneously occurring multiple referents and/or their actions. Thus, while studies on sign language emergence and language development provide an understanding of the emergent trajectory of linearization and the shift from linear to simultaneous constructions expressed by a single articulator, the emergent trajectory of the simultaneous use of multiple articulators to encode distinct semantic elements in an event is not well explored. Only one study provides some insights into the emergent pattern of specific forms of such simultaneity. Namely, in exploring the emergent trajectory of encoding the temporal overlap of events in NSL, Kocab et al. (2016) found a gradual increase in the use of two hands to encode the simultaneity of events. The authors conclude that NSL might be converging on using simultaneous constructions as a linguistic strategy to encode temporal overlap of events, meaning that users can take advantage of the visual modality "in contrast to the strict linearization required by vocal production" (p. 159). Furthermore, the authors argue that the fact that the use of such constructions develops over time indicates that such a device might be challenging in terms of articulation or cognitive load, given that multiple elements must be managed with two hands moving asymmetrically while also controlling for the timing of the manual movement to encode the extent of the overlap of the events (Kocab et al., 2016). However, this study was restricted to the use of both hands to indicate temporality and did not explore the use of simultaneous constructions that recruit not just two but all available articulators to encode distinct semantic elements that are perceptually simultaneous in an event. In the next section we describe how simultaneity and iconicity can be used in a sign language to express such information.

4.1.2. Interplay between simultaneity and iconicity for event encoding in sign languages

Sign languages make use of modality-specific ways to encode perceptually simultaneous events in a simultaneous manner by resorting to iconic means of representation (Cormier, Smith, et al., 2015; Cuxac, 1999, 2000; Napoli & Sutton-Spence, 2010; Perniss, 2007a; Quinto-Pozos, 2007; Risler, 2007). Iconicity does not only refer to the resemblance between a single linguistic form and its meaning, i.e., *imagistic iconicity* (Taub, 2001), but it also refers to the structural resemblance of the relationship between multiple meaning elements and the relationship between multiple elements of linguistic form, i.e., *diagrammatic iconicity* (Perniss, 2007a; Risler, 2007; Slonimska et al., 2021; Taub, 2001).

For the most part simultaneity in sign languages has been investigated by studying the simultaneous use of both hands to encode different events or processes happening at the same time (e.g., Kocab et al., 2016; Napoli & Sutton-Spence, 2010; Risler, 2007). However, signers use not only both manual articulators in a simultaneous manner but also non-manual articulators such as the torso, head, eye gaze direction and facial expression to encode distinct semantic information. Accordingly, signers can vary the information density of a simultaneous construction (i.e., number of simultaneously encoded semantic information units) when encoding one or multiple events and their elements (Dudis, 2004; Slonimska et al., 2020). In a simultaneous construction, signers use imagistic and diagrammatic iconicity to encode a complex event (i.e., consisting of multiple meaning elements and their relationship). Namely, while imagistic iconicity can be used to represent individual meaning elements of the event, diagrammatic iconicity is used to establish a motivated relationship between them. In other words, each element of the event encoded by different articulators (e.g., two hands) can be interpreted in a diagrammatic relationship relative to each other, resulting in a simultaneous representation of an entire event. For example, *depicting constructions*, also known as *classifier* *constructions* (see Schembri, 2003, for an overview of the terminology), where a signer depicts an event with their hands in front of their body (i.e., signing space) on a miniature scale, can be used to encode a plethora of static and motion events involving multiple referents (Perniss, 2007). To encode a motion event, e.g., a horse jumping over a fence, a signer can use one hand to depict the fence and the other to depict the horse (i.e., imagistic iconicity). The arched movement of the hand depicting the horse over the hand depicting the fence represents the horse's jump over the fence (i.e., diagrammatic iconicity). The diagrammatic relationship between both hands reflects the diagrammatic relationship in meaning.

Signers can also make use of the full potential of the visual modality in the form of a highly iconic strategy called *constructed action (CA)*, which employs not only both hands but also the entire upper body to directly map the referent, its actions and emotions onto the different portions of signer's own body, such as face, torso, arms, etc. (Cormier, Smith, et al., 2015; Metzger, 1995). Thus, the event is depicted by the signer at a real-world scale (*character perspective*, Perniss, 2007). As a result of this conceptual mapping, the hands, torso, and head of the signer depicts the respective body parts of the referent, and the emotion expressed by the signer's face depicts the facial expression of the referent (i.e., imagistic iconicity). In terms of the simultaneous expression of meaning, this strategy has the potential to express multiple units of information about the event in a single instance, since each articulator can be recruited to encode a semantic unit of information. For example, in order to communicate a cartoon image in which a big cat is holding and gazing at a small bear, a signer can map the cat onto her own body, using one hand to encode the action of holding while simultaneously using her facial expression and head tilt to mark the cat and its emotion and her eye gaze direction to encode the cat gazing at the bear in its hand (Figure 4-1a). In such an instance, the signer can produce a simultaneous construction containing multiple semantic information units: the cat, the cat's emotion, the cat's eye gaze direction and the action of holding, all of which can be interpreted in a diagrammatic fashion. In order to communicate that the cat not only holds and looks at but also pets the bear, a signer could in addition use her other hand to encode the petting action, i.e., placing one hand above the hand holding the bear and depicting a petting motion (Figure 4-1b). In these examples, representations in each articulator in constructed action are separate units of meaning which can be freely combined with others through diagrammatic iconicity, allowing for a unified interpretation of the event encoded (Slonimska et al., 2021).

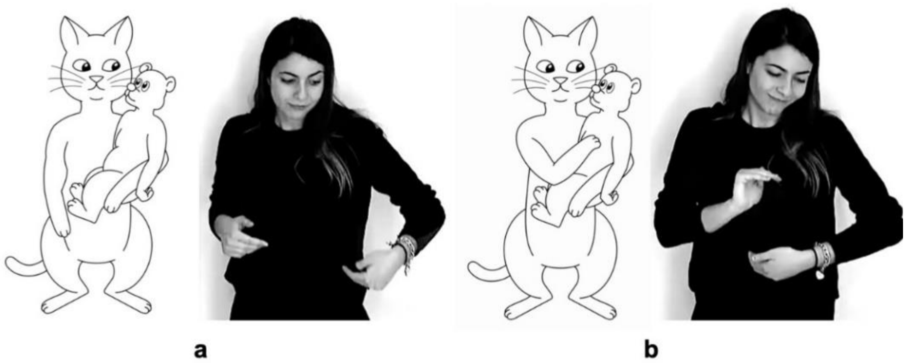


Figure 4-1. A signer encoding a) a cat holding and looking at the bear in its hand, and b) a cat holding, looking at and petting the bear in its hand.

Importantly, considering that meaning elements of the event can be related to each other by means of diagrammatic iconicity in simultaneous constructions, they are not restricted to employing one linguistic strategy, i.e., using only constructed action or only depicting construction. Instead, different linguistic strategies such as constructed action, depicting constructions, lexical signs (i.e., conventionalized manual signs roughly comparable to words in spoken languages), and pointing, can be combined in the same simultaneous construction. For example, in order to encode that the bear is kissing the cat while the cat is holding the bear (Figure 4-2), a signer can use lexical sign for *kissing* by positioning this sign in a diagrammatic relation to the recipient of the kiss, i.e., the cat who is holding the bear (mapped onto the signer's torso, head and eye gaze by means of constructed action).

The use of simultaneous and iconic constructions has been shown to increase as the information load increases indicating that such constructions can be used to achieve communicative efficiency (Slonimska et al., 2020, 2021). We elaborate on this assumption in the next section.

Figure 4-2. A signer encoding a cat holding and looking at the bear in its hand and the bear kissing the cheek of the cat.

4.1.3. The role of simultaneity and iconicity for achieving communicative efficiency in sign languages

Slonimska et al. (2020) showed that signers employ the above mentioned type of simultaneous constructions to cluster related meanings closer together, a phenomenon called *dependency distance minimization* and attested in spoken languages (Gibson et al., 2019; Hawkins, 2004). Dependency distance minimization has been shown to ease information processing in both production and comprehension because it leads to faster syntactic and semantic representation access and as a result boosts efficient communication (Gibson et al., 2019). Efficient communication can be defined as languages being structured "so as to facilitate easy, rapid, and robust communication" (Gibson et al., 2019, p. 389). Communicative efficiency has been studied on all levels of linguistic organization (see Gibson et al., 2019, for an overview), but in our aforementioned study we concentrated on the discourse level, where simultaneity could be potentially used for the same scope as dependency distance minimization in spoken languages, i.e., to cluster related meanings closer together.

In Slonimska et al. (2020) we presented deaf adult signers of LIS with images depicting an event involving animate referents and where the number of semantic information units to be expressed were increased systematically. The participants' task was to describe these images (presented in semi-randomized order) so that the other person (a confederate) could choose the correct image on their laptop. If simultaneity was being used to achieve communicative efficiency, our expectation was that as the amount of information to be communicated increased signers would be likely to increase their use of simultaneous and iconic constructions as well as increase the information density of these constructions (i.e., the number of semantic information units encoded simultaneously) in order to cluster related information as closely as possible. Both predictions were confirmed. We found that signers could freely package multiple information units in simultaneous constructions by also varying their information density. For instance, Fig. 4-1a shows a signer simultaneously encoding two core information units, where the torso, head and eye gaze of the signer represent the cat and the left hand encodes the action of holding, while Fig. 4-1b shows a signer encoding three information units, with the right hand recruited to encode the action of petting. The data showed that signers could encode up to four information units in a single construction at the densest information level by recruiting four different articulators. In the example in Fig. 4-2, the cat is mapped onto the signer's torso, head and eye gaze, while the bear is mapped onto the mouth; the cat's holding action is mapped onto the left hand while the right hand encodes the bear's kissing action.

Furthermore, in a subsequent study that analyzed the same data with regard to the linguistic strategy used to encode the same events, Slonimska et al. (2021) found that the use of constructed action followed a similar pattern: as the amount of information that needed to be encoded increased, so did the use of constructed action. We argued that efficient information encoding is possible due to the fact that constructed action permits articulators from the entire upper body (hands, torso, head, eye gaze, facial expression) to be employed simultaneously to encode distinct semantic information units, thus allowing maximum expressive capacity. This consideration is supported by the finding that signers did not only use constructed action alone but also in combination with other linguistic strategies like lexical signs (e.g., *cat, woman, to kiss* to encode the referent or action), pointing (e.g., to specify the referent) and depicting signs (e.g., entity depicting signs to refer to referent or action). These strategies allowed referents and actions to be placed in the signing space (such as the kissing action in Fig. 4-2) in addition to mapping them onto the body in order to construct informationally dense simultaneous constructions with diagrammatic iconicity.

Together these findings indicate that information organization in sign languages can be achieved through simultaneity, which is made possible due to use of multiple articulators and iconicity. Furthermore, this property can be compared to dependency distance minimization to cluster related meanings closer together and thus achieve communicative efficiency. However, it is also possible that a person who does not know any sign language but is allowed to use only gestures for communication (i.e., silent gesture) can take advantage of the affordances of the visual modality to create such constructions, i.e., using multiple body articulators and iconicity, to simultaneously encode the necessary information units. Therefore, it is crucial to understand whether and to what extent simultaneous and iconic constructions are also available outside of the linguistic system. A systematic comparison of how signers and silent gesturers use simultaneity might shed the light on whether the way simultaneity is used in sign languages constitutes a linguistic resource (i.e. simultaneous compositionality) that has evolved for greater communicative efficiency, that is, expressing together meaning units that are related to each other.

4.2. The present study

In the present study we conducted the same experiment as in Slonimska et al. (2020) with Italian speakers asked to use only their gestures to communicate, in order to assess how silent gesturers express increasing information demands and how their encodings compare to those of signers. We adjusted a few aspects of the elicitation and coding to ensure that silent gesturers could do the task in a way that would be comparable to what LIS signers did.

Slonimska et al. (2020) argued that signers use simultaneous constructions to achieve communicative efficiency by clustering related meanings together and thus reducing dependency distances. These findings showed that when the amount of information to be communicated increased, signers increased the use of simultaneous constructions as well as the information density of these constructions. If the simultaneity employed by signers in Slonimska et al. (2020) reflects a general affordance of the visual modality to encode multiple meaning elements through iconicity, we would expect that silent gesturers would also recruit simultaneity to the same extent as signers or use it even more than signers. However, considering the recent claims about the role of simultaneous and iconic constructions in achieving efficient communication in sign languages as well as research showing the later emergence of simultaneous structures in sign languages, it is possible to hypothesize that sign languages adapt for communicative efficiency through evolution of simultaneous constructions to allow encoding more information as closely together as possible. In this second hypothesis, then, we would expect silent gesturers to use less simultaneity than signers as the amount of information needing encoding increases and to possibly use simultaneity in qualitatively different ways than signers, indicating that it constitutes an emergent linguistic resource for achieving communicative efficiency in sign languages. In the latter scenario, sign languages having already established conventional segmented units and a system for their combination could more easily bring them together allowing simultaneous compositionality in comparison to gesturers who might prefer to express same information through holistic forms that do not allow segmented units to be expressed simultaneously.

4.3. Methods

The study has been approved by the Ethics Council of the National Research Council of Italy (protocol n. 0012633/2019). The method is based on the method developed by Slonimska et al. (2020). We report it below and add the relevant changes in procedure and in coding due to the differences in testing groups.

4.3.1. Participants

Twenty-three hearing Italian adults (12 female, M age = 26.04, range 18–37) participated in the study. All participants were native speakers of Italian with no knowledge of LIS or any other sign language. Participants were recruited via a mailing list made available to The Institute of Cognitive Sciences and Technologies and via an advertisement posted on various social media sites. All participants signed consent forms, agreeing to be video-recorded and giving permission for their data to be used for academic and scientific purposes. Participants received 5 EUR for their participation. This is the same procedure followed in Slonimska et al. (2020) to recruit the 23 LIS signers whose data are also analyzed here.

4.3.2. Design and material

The study was based on the design developed by Slonimska et al. (2020). Items used for data collection are freely available online (https://osf.io/g57p2).

The experimental items consisted of 30 images divided across five levels, with each consecutive level representing an increase in the information density of the event (*Information density levels*). All images (PNG images for levels 1 and 2 and GIFs for levels 3, 4, and 5) depicted two animate referents and their action/s. In total, there were six different referent pairs (e.g., a bird and a bunny) with five information density levels each. At *Information density level 1*, two target information units required encoding—referent 1 and referent 2. At the other end, *Information density level 5*, five target information units required encoding—referent 1, referent 2, static action of referent 1, dynamic action of referent 1, & dynamic action of referent 2 (see Figure 4-3 for all levels and information units).

Five referent pairs depicted animals and one referent pair depicted humans (i.e., a woman and a child). In different referent pairs, the animal referents alternated between referent 1 and referent 2. The following animal referent pairs were created: a dog and a bird, a bird and a bunny, a bunny and a cat, a cat and a bear, a bear and a dog. Referent 1 was always the bigger and referent 2 was always the smaller referent. The items depicted two types of action, static and dynamic, which were considered as additional information units. The static action was the action of referent 1 *holding* referent 2, and it remained constant throughout the items. Dynamic actions were repetitive actions and GIFs were used to capture their motion. The dynamic action of referent 1 *petting* referent 2 was held constant, but the dynamic action of referent 2 varied based on the specific referent pair and was an action that could be performed by the hand/paw (*tapping, petting, pinching*) or by the head/mouth (*pecking, licking, kissing*).

With the increase of information density level, other information units were added in an increasing manner (with the exception of levels 3 and 4, which varied in their perceptual complexity, see below):

- *Information density level 1* (two referents = two information units requiring encoding);
- *Information density level 2* (two referents + one static action = three information units requiring encoding);
- *Information density level 3* (two referents + one static action + one dynamic action by referent 1 = four information units requiring encoding);
- *Information density level 4* (two referents + one static action + one dynamic action by referent 2 = four information units requiring encoding). Note that the information density at this level is the same as in the previous level. The difference between levels 3 and 4 is based on the increase in perceptual complexity: as opposed to one patient and one agent in level 3, in level 4 both referents are patient and agent at the same time;
- *Information density level 5* (two referents + one static action + one dynamic action by referent 1 + one dynamic action by referent 2 = five information units requiring encoding).

The eye gaze of the referents was kept constant and thus both referents were looking at each other throughout. Following Slonimska et al. (2020), semantic information about eye gaze or the size difference between the referents was not in the design and their encoding was not included as separate information units in the analyses, even though participants did sometimes encode it. Participants did not encode any other additional information unit in our data set.

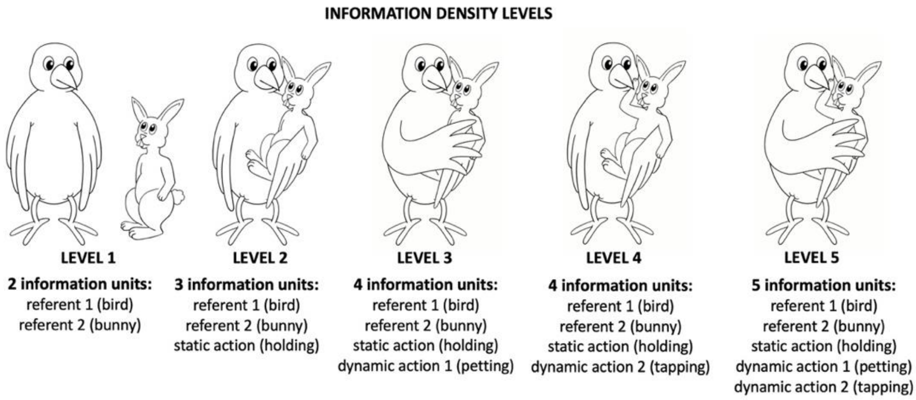


Figure 4-3. PNG items for level 1 (a) and level 2 (b), GIF items for Level 3 (c), level 4 (d) and level 5 (e). In the GIFs, the dynamic action of referent 1 (petting) and dynamic action of referent 2 (tapping) are animated. All original items can be found online (https://osf.io/g57p2).

4.3.3. Procedure

Each participant was told that they would play a *director-matcher* game, where they were assigned the role of *director* and another person was assigned the role of *matcher*. The participant stood in front of the other player, who was seated at a table with a laptop, facing the participant. The director's task was to describe images so the matcher could identify the correct one. The matcher was a confederate who had been instructed to look attentively at what the participant was producing and provide positive feedback (i.e., a head nod, an *OK* gesture or the word *OK*, a thumbs-up gesture) after the participant had finished their description in each trial. No verbal or non-verbal signals of feedback indicating doubt were given in order to ensure that the feedback given to all participants was homogeneous. Participants were informed that the matcher was viewing on their laptop multiple images containing different referents interacting with each other to ensure that participants gave informative descriptions. No additional information regarding what the matcher could see on the laptop was given to the participants. In reality, the confederate had the target image open on the laptop and thus knew which image was being described. To the left side of the table and outside the matcher's view was a 40-inch screen on which the stimuli were presented to the participant. The experimenter was seated at the left end of the table and controlled the presentation of the trials by means of a different laptop connected to the TV screen.

Before the start of the actual game, participants were first asked to look at all the referents that had been picked for the game; they were presented one by one in a PowerPoint presentation. Participants were asked to describe these referents to the matcher using silent gesture. This task functioned as a warm-up and allowed the director the opportunity to describe each referent in detail (e.g., round ears, big paws, whiskers, etc.) and the director and matcher to form a common understanding of the referents. This was also done so that during the presentation of the experimental trials participants would feel less need to concentrate on providing details meant to identify the referent and instead focus on describing the relations and actions between the referents. During the warm-up, once the director had described a referent and the matcher had nodded to indicate they had understood and picked the image on their laptop, the experimenter proceeded to the next image.

Once all referents had been named and participants had no further questions, the actual experiment began. Participants would see an image on the TV screen and describe it to the matcher using silent gesture. If the participant omitted an information unit that required encoding (see Section 4.3.2.), resulting in an incomplete description, the experimenter asked the participant to look at the image carefully once more and repeat the production. However, following Slonimska et al. (2020), the repaired productions were not considered for the analyses in order to assure that all data points analyzed were produced under the same communicative conditions, that is, first descriptions. All trials were presented in a semi-randomized order so that the same referent pair and the same information density level did not appear consecutively. This ensured that in order to do the task participants had to encode all the information units depicted in the images rather than contrasting only specific information units. Thus, each image was always described independently of other images. Participants' productions were video-recorded and the recordings were used for coding. Each participant described 30 images in total but, as mentioned above, only complete descriptions (i.e., first descriptions containing all information units that required encoding) were included in the analyses. Of a total of 690 experimental trials, 43 trials were incomplete. After the experiment ended, the participants were debriefed about the study's procedure and goals.

4.3.4. Coding

The video-recorded data was coded in multimodal data annotation software ELAN, developed by Max Planck Institute for Psycholinguistics (Wittenburg et al., 2006). For the present study, the coding scheme used in (Slonimska et al., 2020) was slightly adapted to accommodate the fact that a) silent gesturers used multiple gestures to refer to a single referent, and b) silent gesturers employed their entire body to encode meaning and they also used non-manual articulators without the simultaneous use of hands (unlike signers, who always used their upper body simultaneously with signs encoded by the hand/s). We elaborate on the differences in coding productions for signers and silent gesturers below. For each stimulus (annotated with its information density level and referent pairing) we coded: *Length of encoding, Kinematic simultaneity*, *Information density of simultaneity.*

4.3.4.1. Movement segments and length of encoding

To determine the linear organization of the production, we followed Slonimska et al. (2020) and segmented data into movement segments (MS) based on the start and end of a movement produced by the participant. A movement segment could be classified as *manual gesture, non-manual gesture*, or *whole body movement*. Movement segments that included manual gesture strokes (with or without use of non-manual articulators and/or body movement) were classified as *manual gestures*, while movement segments produced only by non-manual articulators were classified as *non-manual gestures* (e.g., a gesture stroke performed using only the head and lips (i.e., no hands) to encode the dynamic action *kissing*; see Fig. 4-4, MS5). In addition, silent gesturers used whole body movements, stepping to the right/left space and remaining there to encode referents (see Fig. 4-4, MS3, MS4a, MS4b, MS5). If they were used in absence of manual gestures or non-manual gestures, a movement segment was coded as *whole body movement*, as in Fig. 4-4, MS3 (as opposed to *manual* or *non-*

manual gesture). If the whole body movements were used simultaneously with manual or non-manual gestures, they were classified accordingly – *manual gesture* in the former case (Fig. 4-4, MS4a, MS4b) and *non-manual gesture* in the latter case (Fig. 4-4, MS5). Note that in Slonimska et al. (2020) signers never used non-manual articulators independently from manual signs. Also, signers never used whole body movements involving stepping to the right or left space to distinguish referents. Thus, while for signers all movement segments could be classified as *manual gestures* (which could include non-manuals simultaneously), for silent gesturers, movement segments could be also *non-manual gestures* or just *whole body movement*. Following Slonimska et al. (2020), movement segments could also include the holds of the previous gesture if maintained from one movement segment to the next.

To determine the length of the production for each stimulus, we counted the total number of movement segments produced. Here we observed another difference between signers and silent gesturers: unlike signers, silent gesturers sometimes produced compounded gestures to identify the same referent by listing their multiple features (e.g., a gesture for *whiskers* followed by a gesture for *ears* to refer to *a cat*, see Fig. 4-4 MS1a and MS1b; a gesture for *ears* followed by a gesture for *cheeks* to refer to *a bear*, see Fig. 4-4 MS4a and MS4b). However, in LIS, one sign is always used to identify the referents present in the experiment. Importantly, as we were not interested in how referents are named but rather how event units are constructed and related to each other, we treated compounded gestures for naming referents as a single movement segment in silent gesturers in order to be able to make comparisons to signers. For example, in Fig. 4-4 each of the encodings for *cat* and *bear* consisted of 2 gestures to form a compound name for the referent (MS1a & MS1b for *cat*, and MS4a & MS4b for *bear*). Such identifications of the referents were treated as a single movement segment in the analyses.

Following Slonimska et al. (2020) we excluded all movement segments that were clear disfluencies or mistakes after which gesturers corrected themselves as well as additional movement segments that added extra information that was not the focus of our study (i.e., size or shape of the referents, movement segments encoding only the eye gaze direction of referents).

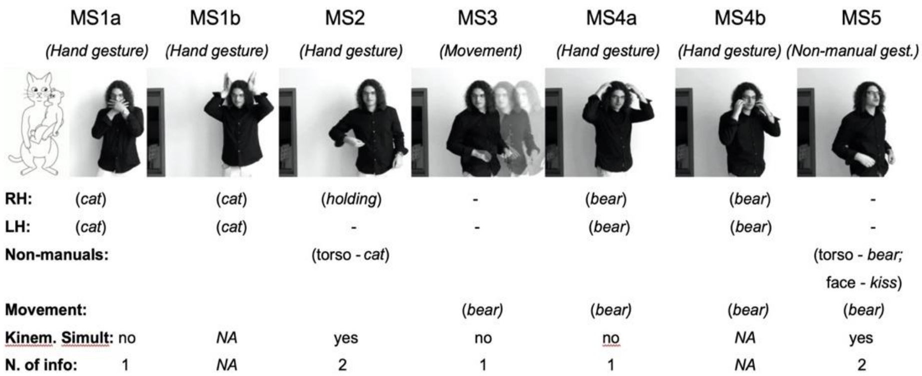


Figure 4-4. Example of a silent gesturer's segmentation of the movement segments of a single stimulus (*Information density level 4*, referent pair: cat - bear): 5 movement segments in total.

4.3.4.2. Kinematic simultaneity

For each movement segment we coded whether articulators (manual and non-manual: left hand, right hand, torso, head, eye gaze, facial expression) and whole body movement (to the right/to the left) were used simultaneously to encode different information units (referent 1, referent 2, static action, dynamic action 1, dynamic action 2). If more than one articulator was used to encode different information units, the movement segment was considered kinematically simultaneous. For example, in MS1a (Fig. 4-4) the gesturer uses both hands to encode the cat – referent 1 (Fig. 4-4, MS1a, RH, LH). No other target information is encoded in this movement segment, so it was coded as not kinematically simultaneous and containing only one information unit. In contrast, in MS5 the whole body movement of the gesturer (Fig. 4-4, MS5, Movement) together with use of the torso encoded referent 2 – the *bear* (see Fig. 4-4, MS5, Non-manuals), while the facial expression, or more specifically the gesturer's mouth, encoded the action of kissing (see Fig. 4-4, MS5, Non-manuals). In MS5 the gesturer's hands are in a resting posture, so assessed based on the gesturer's hand position before starting the production. Accordingly, in MS5, two information units are encoded, which makes this movement segment kinematically simultaneous in our coding. A movement segment containing two or more information units was coded as kinematically simultaneous, reflecting the simultaneous organization of information (more than one information unit in one movement segment) versus the linear organization of information (one information unit in one movement segment). In analyses, kinematic simultaneity was assessed by means of a proportion calculated by dividing the number of movement segments with kinematic simultaneity by the total number of movement segments used per trial.

4.3.4.3. Information density of simultaneity

Finally, we counted how many information units were encoded within each movement segment that contained kinematic simultaneity. In each movement segment, we counted how many information units of interest were simultaneously and explicitly available to the interlocutor (see Fig. 4-4, *N. of info*.). As in Slonimska et al. (2020), if silent gesturers encoded the action of the referent with a hand, the referent itself had to be marked by at least one non-manual marker (eye gaze, facial expression, head, torso) or whole body movement in order for it to be counted as a separate information unit (e.g., Fig. 4-4, MS2). Three information units in a single movement segment could be depicted by, for example, using the torso, head, face and/or an eye gaze direction to encode referent 1, while one of the hands could be used to encode the static action and the other hand could encode dynamic action 1 (Figure 4-5). Four information units could be represented by using the head, facial expression and eye gaze direction to encode referent 1, the hands to encode the static action and dynamic action 1 and the gesturer's torso to encode referent 2, considering that the action of petting was directed to or in contact with the gesturer's torso area, representing the patient of the petting action (Figure 4-6).

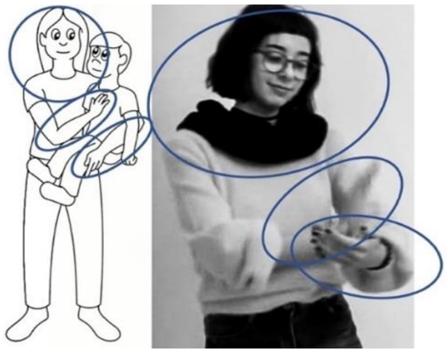


Figure 4-5. An example of a movement segment (MS) with three information units. Circles represent semantic information units of the same referent (the woman).

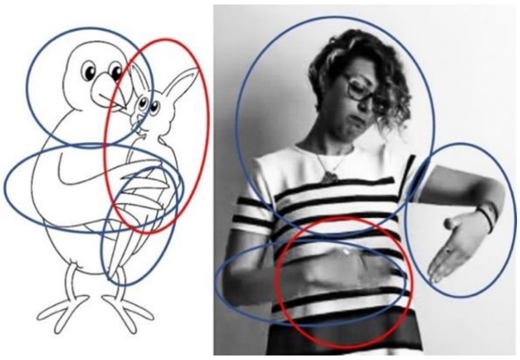


Figure 4-6. An example of a movement segment (MS) with four information units. Circles represent semantic information units of different referents (blue circles for the bird, red circles for the bunny).

4.3.5. Reliability

Data was coded by two trained coders naive to the hypotheses of the study. The first author of the study independently coded 20% of the data.

Agreement between coders was almost perfect – 96.2% for gross-level segmentation of movement segments. Of 812 annotated movement segments, coders agreed on 809 movement segments. We then derived the reliability statistic by assessing the total number of movement segments per response to a trial. Reliability was very strong, as revealed by Cohen's κ of 0.97. Reliability for movement segment type (*manual gesture, non-manual gesture, whole body movement*) between coders was very strong (*Cohen's κ* = 0.89). Reliability for simultaneous use of multiple articulators (manual and non-manual) in each movement segment (*Cohen's κ* = 0.91) as well as reliability regarding number of simultaneously encoded information units in each movement segment were also very strong (*Cohen's κ* = 0.86).

4.4. Results

We first present results from silent gesturers with regard to length of encoding, kinematic simultaneity and information density of simultaneity. We then compare productions by silent gesturers to those by signers in the same order.

4.4.1. Silent gesture

In total we analyzed 647 experimental trials (43 out of 690 trials were excluded due to incomplete descriptions), where production added up to 3145 movement segments used. Of all movement segments, 2625 were classified as *manual gestures*, i.e., a movement where at least one hand was used to perform a stroke, 270 were classified as *non-manual gestures*, i.e., gestures performed with body parts other than hands, and 250 were classified as *whole body movements*, where a participant made a step to the left/to the right and the rest of the articulators were not employed.

The results are organized as follows: we first analyzed the effect of *Information density level* on *length*, *kinematic simultaneity* and *information density of simultaneity* of productions using silent gesture. We then compared the data from the group of silent gesturers to the data collected from the group of signers and analyzed in Slonimska et al. (2020). Following Slonimska et al. (2020), quantitative analyses were performed using generalized mixed models (*lme4* package, Bates et al., 2015). A random structure model was built based on the maximal effects structure that converged (Barr et al., 2013). For all independent variables analyzed, i.e., *length of encoding, simultaneity* and *information density of simultaneity*, the random structure model contained random intercepts for *participant* and *trial*. We then performed model comparisons using ANOVA tests to account for possible confounding factors such as *gender*, *age*, *handedness*. None of these effects improved the baseline models. The final baseline model included random intercept of *participant* and random intercept of *trial* for analyses of *length, simultaneity,* and *information density of simultaneity*. For all models, forward difference coding was used to specify hierarchical contrasts for consecutive information density levels. For a detailed view of the all datasets, R script code and analyses please visit the dedicated OSF repository (https://osf.io/uw2jd/).

4.4.1.1. Length of encoding

The participants tended to increase the length of their productions by using more movement segments as the Information density level increased (level 1: M=3.34, SD=0.60; level 2: M=4.06, SD=0.74; level 3: M=5.11, SD=1.07; level 4: M=5.35, SD=0.89; level 5: M=6.56, SD=1.36). In Figure 4-7, we present the raw means for the total movement segments (MS) used in each trial and for each Information density level in the silent gesture condition tested in the present study. The analysis is based on 647 data points, which represent each experimental trial.

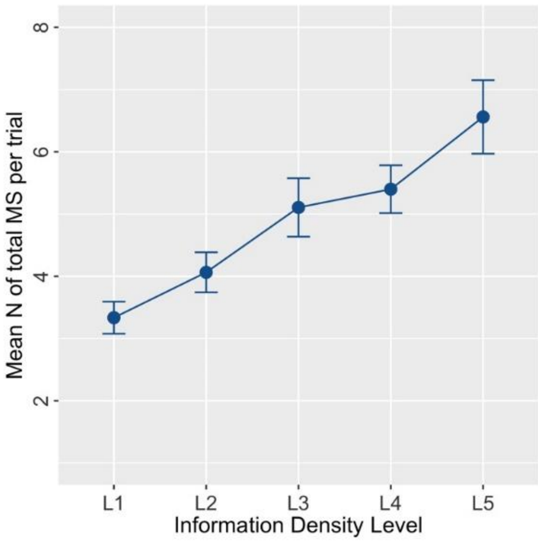


Figure 4-7. Means of total number of movement segments (MS) used per response to a trial by silent gesturers. Error bars indicate a 95% CI for observations grouped within participants.

In order to assess whether the increase in length in the silent gesture group was statistically significant, we fitted a Poisson mixed effects model. The following possible confounding factors were accounted for: *gender* ($\chi2(1) = 1.54$, $p = .214$)*, age* ($\chi2(1) = 0.79$, $p =.374$) and *handedness* ($\chi2(1) = 0.004$, $p = .950$). None of the factors was significant. The model including *Information density level* (categorical variable with 5 levels) was compared to the baseline model which included the random intercept for *participant* and the random intercept for *trial*. The dependent variable was the *length of the production*, quantified as the total number of movement segments used per trial. We found that the model including the fixed effect of *Information density level* improved model fit over the baseline model ($\chi2(4) = 70.41$, $p < .001$, see Table 4-1 for a summary of the model). The contrasts between levels shown in Table 4-1 indicate a significant gradual increase in length of productions, except levels 3 and 4, which were comparable.

Random effects	**Variance**	**SD**				
Trial	0.00[11]	0.00				
Participant	0.008	0.088				
Number of obs: 647	Group: Trial=30	Participant=23				
Fixed effects	**β**	**95% CI** Lower b.	Upper b.	**SE**	***z***	***p***
(Intercept)	1.56	1.51	1.61	0.03	60.24	<.001
Level 1 vs. Level 2	-0.20	-0.32	-0.07	0.06	-3.10	.002
Level 2 vs. Level 3	-0.23	-0.34	-0.11	0.06	-3.94	<.001
Level 3 vs. Level 4	-0.07	-0.17	0.04	0.05	-1.20	.23
Level 4 vs. Level 5	-0.19	-0.29	-0.08	0.05	-3.55	<.001

Table 4-1. Best fit model on a logit scale (model fit by maximum likelihood, Laplace approximation) regarding use of total number of movement segments per experimental trial. Contrasts reflect comparisons between levels in hierarchical order attained through forward difference coding.

[11] The zero variance for the random effect of trial is driven by the inclusion of the fixed effect of *Information density level*, which accounts for all variance detected in random effect of trial in the baseline model. Given that the inclusion of trial is based on the initial design of the study and the results do not change if this random effect is left out, we have kept it in the primary model. Controls of the random effect of trial can be found in the supporting material. This consideration applies to all consecutive analyses.

4.4.1.2. Kinematic simultaneity

The analysis of kinematic simultaneity was based on the mean proportions of the movement segments with kinematically simultaneous articulators (i.e., two or more articulators used in a single movement segment) expressing distinct information out of the total number of movement segments used per trial (Figure 4-8). In level 1, use of simultaneity was close to non-existent (M=0.007, SD=0.02) though it increased consecutively in level 2 (M=0.21, SD=0.10), level 3 (M=0.36, SD=0.09), level 4 (M=0.41, SD=12), and level 5 (M=0.46, SD=0.12).

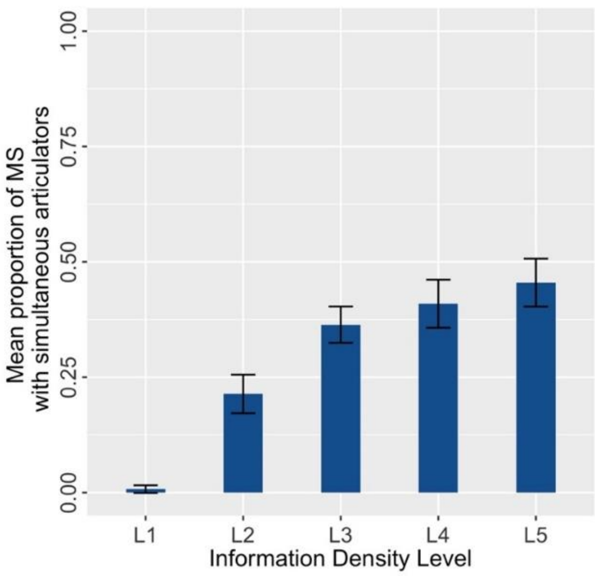


Figure 4-8. Mean proportions of kinematically simultaneous movement segments (MS) out of total number of movement segments per trial. Error bars indicate 95% CI of observations grouped within participants.

To assess whether the increase of simultaneity was significant, a logistic mixed effects model was fitted to assess the effect of *Information density level* on *kinematic simultaneity*. Possible confounding factors *gender* ($\chi2(1) = 0.86$, $p = .35$)*, age* ($\chi2(1) = 0.30$, $p = .59$)*,* and *handedness* ($\chi2(1) = 1.02$, $p = .31$) were not significant. We compared the model containing *Information density level* to the baseline model which contained random intercepts of *participant* and *trial* and found a significant improvement in the model ($\chi2(4) = 122.37$, p <.001, see Table 4-2 for a summary of the model). The contrasts between levels shown in Table 4-2 indicate a significant gradual increase in kinematic simultaneity, except levels 3 and 4, which were comparable.

Random effects	Variance	SD			
Trial	0.00	0.00			
Participant	0.15	0.39			
Number of obs: 647	Group: Trial=30	Participant=23			
Fixed effects		**95% CI**	**SE**	***z***	***p***
	β	Lower b.	Upper b.			
(Intercept)	-1.46	-1.72	-1.20	0.13	-10.95	<.001
Level 1 vs. Level 2	-3.45	-4.45	-2.44	0.51	-6.72	<.001
Level 2 vs. Level 3	-0.75	-1.01	-0.49	0.13	-5.60	<.001
Level 3 vs. Level 4	-0.14	-0.37	0.08	0.11	-1.25	.21
Level 4 vs. Level 5	-0.24	-0.45	-0.02	0.11	-2.18	.03

Table 4-2. Best fit model on a logit scale (model fit by maximum likelihood, Laplace approximation) regarding the proportion of kinematically simultaneous movement segments. Contrasts reflect comparisons between levels in hierarchical order attained through forward difference coding.

4.4.1.3. Information density of simultaneity

Overall, 3145 movement segments were produced. In total, 2149 movement segments encoded only one information unit, 727 movement segments encoded two information units, 262 movement segments encoded three information units and only 7 movement segments encoded four information units (Figure 4-9). However, 6 of the 7 movement segments containing four information units were produced by a single participant, and the remaining movement segment was produced by another participant. The analyses are based on 3145 data points representing each movement segment.

In order to assess whether *information density of simultaneity* increased as the *Information density level* increased, we fitted a Poisson mixed model. Possible confounding factors *gender* ($\chi2(1) = 0.96$, $p = .32$)*, age* ($\chi2(1) =$ 8e-04, $p = .98$) and *handedness* ($\chi2(1) = 1.02$, $p = .31$) were not significant. We compared the model containing *Information density level* to a baseline model containing random intercepts of *trial* and *participant*. We found that the model including the fixed effect of

Information density level improved model fit over the baseline model ($\chi2(4) = 67.19$, $p<.001$, see Table 4-3 for a summary of the model). The contrasts between levels shown in Table 4-3 indicate a significant gradual increase in information density of simultaneity, except levels 3 and 4, which were comparable.

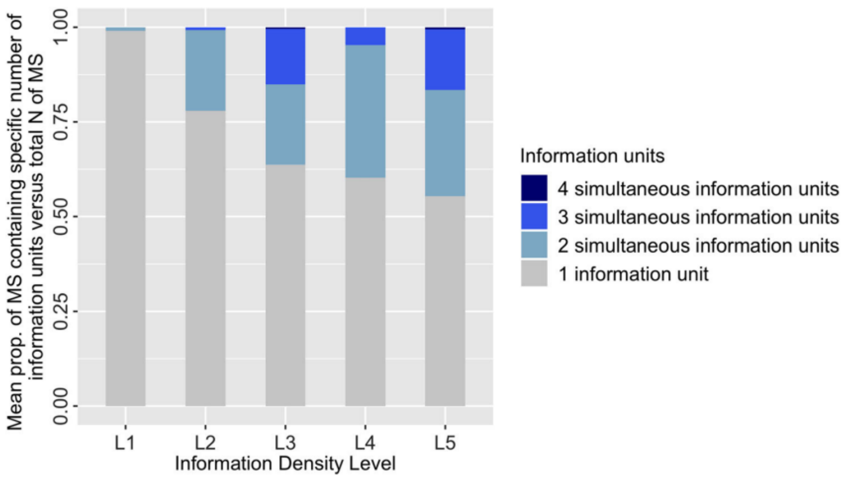


Figure 4-9. Mean proportions of movement segments (MS) with 1, 2, 3, and 4 simultaneous information units out of the total number of movement segments in the silent gesture group

Random effects	**Variance**		**SD**			
Trial	0.00		0.00			
Participant	0.002		0.05			
Number of obs: 3145	Group: Trial=30		Participant=23			
Fixed effects		**95% CI**		**SE**	***z***	***p***
	β	Lower b.	Upper b.			
(Intercept)	0.30	0.26	0.33	0.02	15.54	<.001
Level 1 vs Level 2	-0.19	-0.31	-0.07	0.06	-3.16	.002
Level 2 vs Level 3	-0.22	-0.31	-0.12	0.05	-4.31	<.001
Level 3 vs Level 4	0.05	-0.04	0.14	0.04	1.13	.26
Level 4 vs Level 5	-0.12	-0.20	-0.03	0.04	-2.73	.006

Table 4-3. Best fit model on a logit scale (model fit by maximum likelihood, Laplace approximation) regarding the increasing information density of simultaneity. Contrasts reflect comparisons between levels in hierarchical order attained through forward difference coding.

4.4.2. Silent gesture vs. sign language

In order to compare productions in silent gesture and in sign language, we used the data collected from 23 participants in the present silent gesture study and combined them with the data from 23 deaf signers of LIS for the same task and collected by Slonimska et al. (2020). The length and kinematic simultaneity comparisons are based on 1325 trials in total and the information density of simultaneity comparison is based on 6842 movement segments in total. We first report quantitative analyses of the length of encoding followed by quantitative and qualitative analyses of kinematic simultaneity and information density of simultaneity.

4.4.2.1. Length of encoding

We ran a series of Poisson mixed models in order to assess the effect of *group* (silent gesture or sign) on the length of the productions (Table 4-4). We first compared a model containing the fixed effect of *Information density level* to a baseline model which included random intercepts for *participant* and *trial*. The model was significantly improved by including this variable ($\chi2(4) = 95.76$, $p<.001$). We then added a model including the fixed effect of *Group* and then the interaction between *Information density level and Group*. The model was improved by adding the fixed effect of *Group* ($\chi2(1) = 5.12$, $p=.024$) but not the interaction ($\chi2(4) = 2.16$, $p=.71$). The *Group* effect revealed that overall, the productions by silent gesturers were significantly shorter (β=-0.10, CI [-0.19; -0.02], SE= 0.04, $z=-2.33$, $p=.02$) than the productions by signers (Figure 4-10).

Models	**Df**	**AIC**	**BIC**	logLik	devianc	χ2	**Df**	***p***
Baseline	3	5234.8	5250.4	-2614.4	5228.8			
Info.density	7	5147.0	5183.4	-2566.5	5133.0	95.76	4	**<.001**
Group	8	5143.9	5185.4	-2564.0	5127.9	5.12	1	**.024**
Info.density x Group	12	5149.8	5212.0	2562.9	5125.8	2.16	4	.706

Table 4-4. ANOVA model comparisons of the fixed effects of *Information density level*, *Group* and interaction between *Information density level* and *Group* for the length of the productions.

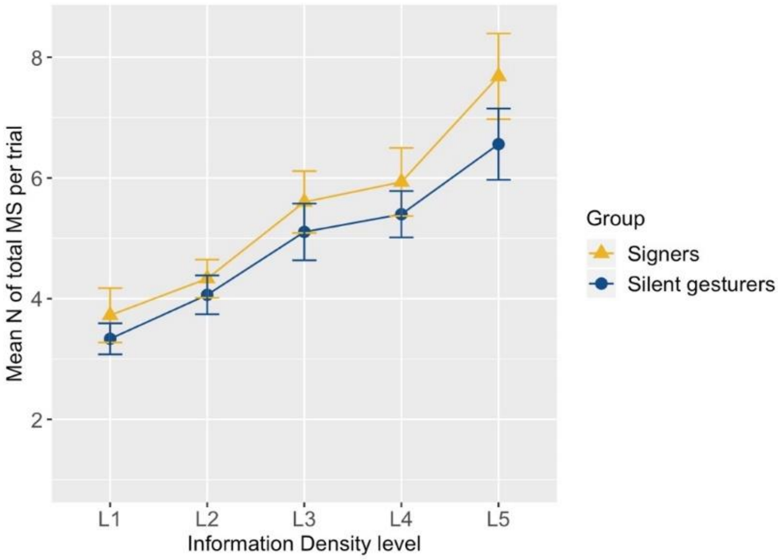


Figure 4-10. Means of total number of movement segments (MS) used per each trial for silent gesture and sign language groups. Error bars indicate 95% CI of observations grouped within participants.

4.4.2.2. Kinematic simultaneity

We ran a series of logistic mixed models in order to assess the effect of *Group* (silent gesture/ sign language) on the kinematic simultaneity (Table 4-5). We first compared a model containing a fixed effect of *Information density level* to a baseline model which included random intercepts of *participant* and *trial*. We found that the model was improved significantly by this factor ($\chi2(4) = 140.57$, $p<.001$). We added a model containing a fixed effect of *Group* and then a model with the interaction between *Information density level* and *Group*. The best fit model included the fixed effect of *Group* ($\chi2(1) = 37.83$, $p<.001$) while the addition of the interaction did not reach the level of significance (see Table 4-5). The *Group* effect revealed that overall the productions by silent gesturers were significantly less simultaneous (β=-0.98, CI [-1.23; -0.73], SE= 0.13, z= -7.65, $p<.001$) than the productions by signers (Figure 4-11).

Models	Df	AIC	BIC	logLik	deviance	χ2	Df	*p*
Baseline	3	3078.5	3094.1	-1536.3	3072.5			
Info.density	7	2946.0	2982.3	-1466.0	2932.0	140.57	4	**<.001**
Group	8	2910.1	2951.7	-1447.1	2894.1	37.83	1	**<.001**
Info.density x Group	12	2908.8	2971.1	-1442.4	2884.8	9.33	4	.053

Table 4-5. ANOVA model comparisons of the fixed effects of *Information density level*, *Group* and interaction between *Information density level* and *Group* on kinematic simultaneity.

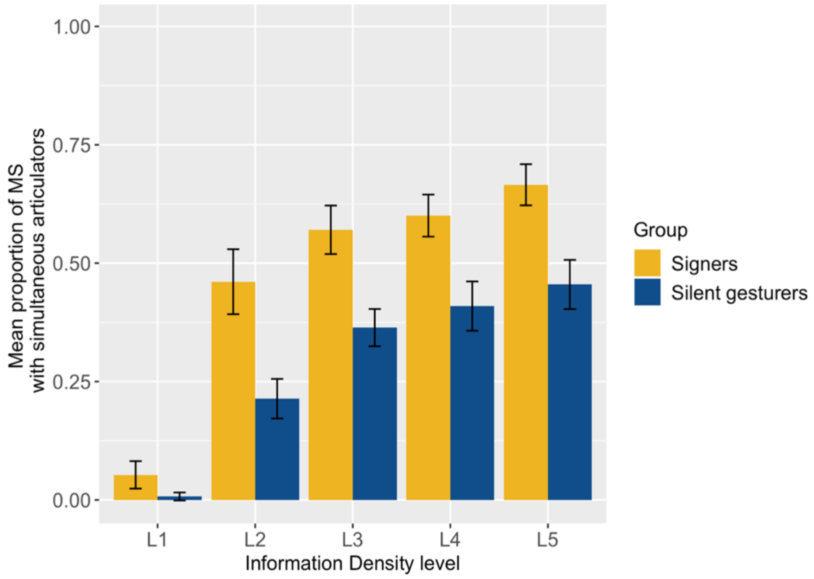


Figure 4-11. Mean proportions of kinematically simultaneous movement segments (MS) out of total number of movement segments per trial for silent gesture and sign language groups. Error bars indicate 95% CI of observations grouped within participants.

4.4.2.3. Information density of simultaneity

We ran a series of logistic mixed models in order to assess the effect of *Group* (silent gesture/ sign language) on information density of simultaneity (Table 4-6). We first compared a model containing the fixed effect of *Information density level* to a baseline model which included the random intercepts of *participant* and *trial*. Including *Information density level* significantly improved the model ($\chi2(4) = 102.69$, $p<.001$). We then added a model with the fixed effect of *Group* followed by a model with the interaction between *Information density level and Group*. The best fit model included the fixed effect of *Group* ($\chi 2(1) = 5.112$, p=.024), while the addition of the interaction did not improve the model significantly. The *Group* effect revealed that overall the productions by silent gesturers were significantly less simultaneously dense (β=-0.21, CI [-0.26; -0.17], SE= 0.02, z=-9.16, p<.001) than productions by signers (Figure 4-12).

Models	Df	AIC	BIC	logLik	deviance	χ2	Df	*p*
Baseline	3	17928	17948	-8961.0	17922			
Info.density	7	17833	17881	-8909.7	17819	102.6	4	**<.001**
Group	8	17788	17842	-8885.7	17772	47.91	1	**<.001**
Info.density x Group	12	17787	17869	-8881.4	17763	8.59	4	.072

Table 4-6. ANOVA model comparisons of the fixed effects of *Information density level*, *Group* and interaction of *Information density level and Group* on information density of simultaneity.

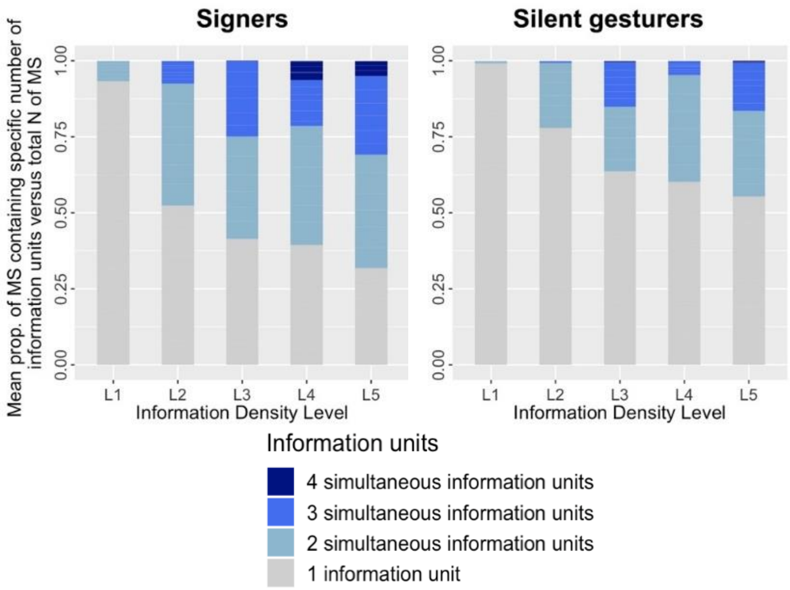


Figure 4-12. Mean proportions of movement segments (MS) with 1, 2, 3, and 4 simultaneous information units out of the total number of movement segments in the sign language and silent gesture groups.

4.4.2.4. Qualitative analysis: differences in simultaneity and length between silent gesturers and signers

Quantitative assessment of the data shows that relative to silent gesturers, signers not only produced longer encodings but they also used kinematic simultaneity as well as information density of simultaneity at higher rates as the number of information units requiring encoding increases. We analyzed the differences in the quality of the simultaneous constructions between groups to see what factors allowed signers to use more simultaneity. For example, we observed that when the dynamic action of referent 2 performed by head/mouth of the referent (i.e., kissing, licking, pecking) needed to be encoded, silent gesturers prototypically directly mapped the referent onto their full body and accordingly performed the kissing/licking/pecking action with the corresponding body part – head/mouth (see Figure 4-13), limiting their expressive possibilities. In contrast, signers always performed such actions with their hands, i.e., mapping mouth-related movements onto their hands, with or without also mapping the action on the mouth (see Figure 4-14). Representing the dynamic action of referent 2 with the hand (via use of lexical signs or the depicting constructions) allowed signers to encode that action together with explicit marking for referent 1 (the patient of the aforementioned action), as well as maintaining the static action (i.e., holding) of referent 1 in a single simultaneous construction (Fig. 4-14 b, d, f), while no such constructions were possible if a direct one-to-one mapping between the head/mouth of the referent 2 and signer was used to represent dynamic action 2 – the prevalent strategy used by silent gesturers. To assess this observation, we counted how many items containing dynamic action 2, performed by head/mouth of the referent, were encoded by using the hand in both groups (see Table 4-7).

There were total of 132 observations for signers and 121 for silent gesturers in levels 4 and 5, which contain the items in which the dynamic action of referent 2 is performed by a head/mouth: kissing, licking, pecking. All signers encoded these actions with their hand. As for the gesturers, of the 23 participants, only 8 used their hand to encode the action of pecking; the majority used their head/mouth (direct mapping). Only 3 silent gesturers used their hand to encode the action of licking (1 of the 8 who also encoded pecking) while none of the silent gesturers encoded the action of kissing with their hand.

Type of action	Action encoded by hand					
	Signers			Silent gesturers		
	Participants (n=23)	N observations		Participants (n=23)	N observations	
Kissing	23	44 (44)	100%	0	0 (38)	0%
Licking	23	46 (46)	100%	2	3 (39)	7.69%
Pecking	23	42 (42)	100%	8	14 (44)	31.81%
Total	**23**	**132 (132)**	**100%**	**9**	**17(121)**	**14.05%**

Table 4-7. Overall number of signers and silent gesturers using a hand to encode an action performed by a mouth/head in an experimental item.

However, in some instances direct mapping of the action could be taken advantage of, i.e., when the dynamic action of referent 2 was performed by a hand (petting, tapping, pinching). More specifically, silent gesturers would use their own hand to encode the action performed by the hand/paw of the referent (Figure 4-15), which allowed them to split this action away from the body representing referent 1 to represent the dynamic action of referent 2 (Fig. 4-15b). This strategy was used in 42% of the observations. Such constructions were somewhat similar to complex constructions used by signers who used them in 64% of the observations (Figure 4-16b).

However, even when the dynamic action of referent 2 was used in constructions where referent 1 was marked on the body, silent gesturers were not likely to take the advantage of their other hand to encode the static action of the referent 1 (12% of observations). Signers, on the other hand, were more likely (72% of observations) to integrate the static action encoded with one of the hands when encoding dynamic action 2 with the other hand and referent 1 with non-manual articulators, thus increasing the number of simultaneously encoded information units in a single movement segment.

pet a b

Figure 4-15. Strategies used by silent gesturers to encode dynamic action 2, performed by the hand of referent 2 (a - prototypical encoding; b - non-prototypical encoding).

pet a b

Figure 4-16. Prototypical strategies used by signers to encode dynamic action 2, performed by the hand of referent 2.

These data show that silent gesturers were mostly constrained by imagistic, one-to-one mapping of the information units (e.g., action of the mouth/head mapped onto the mouth/head of the gesturer), which limited their ability to create more simultaneous constructions and manipulate information density. Signers, on the other hand, could take advantage of the linguistic possibility of encoding the action of the referent by using their hand regardless of the type of action (by means of lexical signs or depicting constructions) and using diagrammatic iconicity to interrelate multiple articulators, thereby allowing more possibilities to construct simultaneous constructions of various degrees of information density.

Furthermore, the fact that signers' encodings were also longer than gesturers' as well more simultaneous and informationally dense in their simultaneous constructions indicates that signers were more redundant. Recall that specific information units that were required to be encoded at each information density level and only these units were analyzed. Accordingly, we can deduce that if participants used more than the specific number of information units in each level it means that some of the information has been encoded more than once, leading to redundancy. For example, in level 3, which required the encoding of four information units, signers used on average 5.60 movements segments, of which 57% were simultaneous constructions. In contrast, at the same level silent gesturers used on average 5.11 movement segments and 36% contained more than one information unit. We also know that signers used more informationally dense simultaneous constructions. Accordingly, we can deduce that signers repeated information units more than gesturers did. While the assessment of the specific information units encoded simultaneously goes beyond the focus of the present study, we can nevertheless conclude that signers are more redundant in their encodings than silent gesturers are. Silent gesturers' less frequent use of redundancy indicates that while they were able to encode all required information units, they were less capable of interrelating these units in diverse simultaneous constructions (Figure 4-17). Signers, on the other hand, did not only encode the necessary information but they also linked meaning units to each other so that each consecutive unit was integrated with the preceding one (Figure 4-18). To do this, signers made use of simultaneous constructions such as pointing with one hand to the representation displayed by the other, thus establishing spatial relation between information units. For example, signers always specified referent 2 not by simply naming it (Fig. 4-18 MS4) but also by deictically referring to it by means of pointing (Fig. 4-18 MS3), establishing a spatial relationship to the static action of holding by referent 1. Silent gesturers did not have the same capacity to establish spatial relationships between referents and their actions, resulting in shorter and more ambiguous productions. Signers, on the other hand, appeared to pay particular attention to encoding the precise relationship between information units. Specifically, already encoded information units could also be maintained in consecutive movement segments encoding new information, leading to greater redundancy. Such an approach enabled signers to chain unfolding information in longer but arguable more coherent structures relative to silent gesturers.

Figure 4-17. Prototypical encoding sequence for *Information density level 4* from gesturers (Referent pair: bird and bunny).

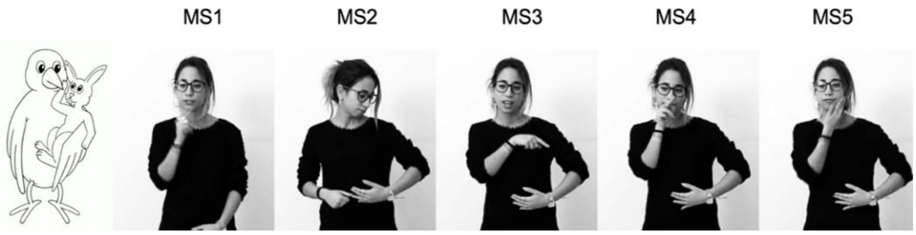


Figure 4-18. Prototypical encoding sequence for *Information density level 4* from signers (Referent pair: bird and bunny).

4.5. Discussion

In the field of language emergence, linearization of the segmented elements is seen as one of the emergent properties of language. Whether emergence of simultaneous constructions composed of multiple meaning units can be also seen as an emergent property arising due to pressures of communicative efficiency has not been investigated. In the present study, we hypothesized that if simultaneity has emerged as a linguistic property for achieving greater communicative efficiency, we would observe that signers use simultaneity more than silent gesturers when faced with increasing information demands. Alternatively, we hypothesized that if simultaneity reflects a general expressive ability to use the visual modality also without having an established linguistic system, silent gesturers would be just as simultaneous in their productions as signers, or even more so. To test these hypotheses, we compared the data from silent gesturers who had no knowledge of any sign language to data collected via the same experimental task from deaf adults who use LIS daily as their main language (Slonimska et al., 2020).

We found that as the amount of information requiring encoding increased, both gesturers and signers increased their use of simultaneity, the density of the simultaneously encoded information and the length of their productions. However, the results revealed that relative to the signers, the gesturers produced shorter encodings, used less kinematic simultaneity and were less informationally dense in the simultaneous constructions they used. Our qualitative analysis showed that linguistic tools (such as the availability of lexical and depicting signs that allow a signer to use a hand to encode an action by another body part) give signers a possibility to construct more simultaneous and more informationally dense constructions. Furthermore, qualitative analysis indicated that the longer and more simultaneous encodings by signers were potentially driven by them being more attentive to the encoding of spatial relationships between referents and their actions and as a result being less ambiguous with regard to who did what to whom. Specifically, signers but not gesturers were likely to maintain specific information units in consecutive movement segments, thus chaining unfolding information into a coherent sequence.

Overall, the results of the present study suggest that the visual modality does afford simultaneous use of more bodily articulators as the need to communicate efficiently increases, something both gesturers and signers did. However, differences between signers and gesturers show that the affordances of modality alone are not enough and a linguistic system is needed to take advantage of this property.

4.5.1. Achieving simultaneity in sign language vs. silent gesture

We hypothesized that if simultaneity is not simply an affordance of the visual modality but instead constitutes a linguistic resource at the disposal of sign language users, then silent gesturers should be likely to use less simultaneity compared to signers. Indeed, results confirmed that it was the case. To put this into perspective, in

Information density level 5, the most informationally dense level of the design, kinematically simultaneous movement segments constituted on average 46% of all productions for silent gesturers. The same percentage of simultaneity was already used at *Information density level 2* in the group of signers, which then accounted for 67% of kinematically simultaneous constructions at *Information density level 5*. Furthermore, three and four simultaneously encoded information units were used considerably more frequently by signers than by gesturers. For example, four simultaneously encoded information units were used only in 7 movement segments in silent gesturers (of which 6 movement segments were produced by the same participant) compared to 102 movement segments for signers. Thus, even though the affordances of the visual modality to use iconicity and multiple articulators for encoding meaning are freely available to both groups, silent gesturers appear not to take full advantage of these properties.

Our findings are in line with Senghas & Littman (2004) who found more simultaneity for encoding motion verbs in an established sign language such as LSE compared to younger cohorts of NSL as well as findings of Kocab et al., (2016) that showed an emergent trajectory of gradual increase of using simultaneity (i.e., two hands) to encode the temporal overlap of the events in NSL. Our data shows that increase in the use of simultaneity is not limited to simultaneously encoded semantic meanings in one articulator or the use of two articulators, but it can be extended to the increase in the use of simultaneity, taking advantage of the full spectrum of available body articulators. Overall, the differences between both groups can be attributed to holistic use of the body and use of imagistic iconicity by silent gesturers versus partitioned use of the body into independent articulators and diagrammatic iconicity by signers. We describe these differences in more detail below.

Qualitative results showed that silent gesturers were more limited in their simultaneous constructions considering that they were more likely to represent referents (referent 1 and referent 2) individually in separate movement segments (see Fig. 4-13 & 4-17). Thus, they did not generally split the action of the referent from the referent itself to integrate it with other information units through diagrammatic iconicity to increase information density of simultaneous constructions. Rather, silent gesturers constructed their messages as a sequence of holistic representations of the referents through imagistic iconicity. This interpretation is also in line with Coppola & So (2006), who argue that silent gesturers produce "pictorial, imagistic representations of events" (p. 128) considering that they lack both lexical gestural repertoire as well as experience with only gestural communication. Indeed, silent gesturers used full enactment of the referent or "acting out", a strategy available not only for deaf but also hearing children even before age of 3 in their gestures (Casey, 2003; Cormier, Smith, & Sevcikova, 2013; Loew, 1984; McNeill, 1996). Namely, imagistically iconic representation inhibited silent gesturers to use more simultaneity considering that they were restricted to one-to-one mapping between body of the signer and body of the referent to encode actions, e.g., in order to represent a kissing action silent gesturers would use their mouth (see Fig. 4-13) rather than using hand to represent kissing as preferred by signers (as in Fig. 4-14). Unlike silent gesturers, signers were not restricted to representing a single referent and its actions only, but they could also encode multiple referents and actions of different referents and by employing variety of linguistic strategies and their combination (constructed action, depicting constructions, lexical signs and pointing) in the same movement segment (Slonimska et al., 2021). More flexibility in regard to how information can be encoded (e.g., encoding kissing action by hand) allowed signers to recruit more diagrammatic iconicity as a structuring resource to relate multiple information units in a single simultaneous construction to represent events more completely. This indicates that simultaneous constructions of signers were a result of simultaneous compositionality through partitioned use of the body as independent articulators and diagrammatic iconicity. In turn, simultaneous constructions produced by silent gesturers were unlikely to be conceptualized as a combination of independent meaning elements but rather as a general affordance to represent the referent's body holistically, e.g., the actions of holding and petting at the same time, even though they involve two hands (i.e., kinematic simultaneity), can be conceptualized as a single body holding and petting.

Taken together, these observations suggest an emergent trajectory of iconic forms, from representing individual entities and their actions holistically through imagistic iconicity to systematically representing whole expressions through diagrammatic iconicity in simultaneous constructions. Thus, during language emergence sign languages might be developing new ways for recruiting iconicity as part of a linguistic system for maximizing the expressive capacity of simultaneous constructions.

4.5.2. Simultaneity as an emergent property of efficient communication in sign languages

While previous research has concentrated on how compositionality emerges through linearization, here we showed that the pressure for communicative efficiency will eventually push sign languages towards simultaneous combination of related meaning units. Crucially, in the present study, we showed that simultaneous constructions that are based on diagrammatic iconicity used by signers differ fundamentally from simultaneity arising from imagistic iconicity present in silent gesturers' productions. Imagistic representations can be useful as a starting point in a communicative system as seen in silent gesturers because they can directly map properties of the entities and actions without needing to have any established conventions. For example, a kissing action performed with the mouth is a direct representation of the action and as such is likely to be unambiguous and therefore communicatively effective. Conversely, the signers have established conventions in the form of manual signs, which allow conventional representations of individual meaning elements that can then be recombined in diagrammatically iconic simultaneous constructions with the rest of the body, allowing them to be both communicatively effective and communicatively efficient. Having a linguistic structure to organize multiple information units simultaneously allows clustering related meanings closer together (i.e., dependency distance minimization) for the benefit of the producer and the perceiver during message production and comprehension. Thus, sign languages evolve the use of simultaneity for achieving greater communicative efficiency.

A question that arises is why do children initially show preference for linear structures if simultaneous constructions are more communicatively efficient? The answer might lie in the fact that before simultaneous constructions can be used, the units of which they are constructed must first emerge in children's language repertoire - as shown for motion verbs (Newport, 1981, 1988; Supalla, 1982) and complex verbs (Meier, 1987; Morgan et al., 2002). Accordingly, during language acquisition, linearized structures are preferred by children as they might facilitate learning of the segmented elements. Once these elements and different linguistic strategies of how they can be encoded are acquired, pressure for communicative efficiency will push towards combining these elements simultaneously. Future research is needed to test this hypothesis. Possibly, the same explanation can be also attributed not only to language development but language emergence in general. Namely, as the linguistic structure emerges it might prioritize linear organization as it aids segmentation and conventionalization of these segmented elements and eventual emergence of compositional structure (Özyürek et al., 2015; Senghas et al., 2004). Once these elements that can be re-used and recombined with other units have been robustly established in a linguistic system ensuring effective language transmission to new learners, simultaneous compositionality can emerge as an adaptation for communicative efficiency during language use. The emergent trajectory of simultaneous constructions constitutes an endeavor for future research that can be potentially assessed experimentally and "in the wild".

4.5.3. Implications for length of encoding and efficient communication

The fact that silent gesturers' productions contained less simultaneity than signers' productions did may suggest that gesturers' productions were also longer than signers' productions. However, that was not the case. In fact, signers' productions not only contained more simultaneity but were also longer than the productions by silent gesturers, indicating higher redundancy in the signers' linguistic signal. These findings contrast with the assumption that communicative efficiency strives for shorter and less redundant structures (Fay, Ellison, et al., 2014; Fay et al., 2013; Kirby et al., 2015; Motamedi et al., 2019). Below we review how co-referential structures – which lead to redundancy of encoded information – can be interpreted in light of communicative efficiency.

The fact that signers' productions were not only more simultaneous and denser but also longer implies that signers repeated necessary information units more than gesturers did. In other words, signers were more redundant than silent gesturers. Redundancy is also a prevalent trait in spoken languages (Hsia, 1977). Arguably, redundancy does lead to greater effort and cost for information transmission, which at first glance might appear to be a property pulling speakers away from communicative efficiency when only sheer length is considered (Motamedi et al., 2019). However, in the present study communicative efficiency is not conceived simply as a reduction in effort and in the cost of information transmission—the notion includes informative quality as well (Futrell et al., 2020; Gibson et al., 2019). Redundancy in languages is used for multiple purposes, including reducing errors in the encoding and the decoding process, reducing the loss of information in noise, interference and distortion, as well as facilitating "association and discrimination, establishing memory traces, and helping to prevent forgetting" (Hsia, 1977, p. 78). As such, redundancy constitutes a crucial factor in achieving communicative efficiency, since it ensures the minimization of information loss and an increase in disambiguation (Fedzechkina et al., 2012; Shannon, 1948). In other words, encoded information is reinforced, thereby maximizing successful comprehension**.** We could speculate that signers used more redundancy to create cohesive productions through anaphorical linkages (Carrigan, 2016). The fact that redundancy in sign languages can be recruited in a simultaneous manner, on the other hand, lessens the strain of effort and cost because it is not necessary to add redundant units in a string; instead they can be incorporated with new encoded information and thus reinforce the relationships between multiple encoded information units (e.g., Fig. 4-18). In systems that rely heavily on the processing of visual information like sign languages, simultaneity can become prioritized over linearity for linguistic organization (Emmorey, 2016; Siple & Fischer, 1991). As such, incorporating redundancy by using simultaneity to add new information might serve to achieve referential cohesion in the unfolding message (Gernsbacher, 1997). In this respect, redundant information functions as an explicit binding element between information that is already encoded and the new information that is being encoding at a given moment in time. In this way, related meanings are encoded closer together, leading to a minimization in dependency distances and greater communicative efficiency.

4.6. Conclusion

Once we keep in mind that in sign languages the visual modality takes the full burden of communicative load, as often pointed out by sign language scholars (Goldin-Meadow et al., 1996; Perniss et al., 2015), it is not difficult to recognize that the unique properties of this modality would be recruited for linguistic organization and communicative efficiency. This study shows that simultaneity is one of these properties and that it emerges from a simple affordance and becomes a linguistic property. Simultaneity is also available to silent gesturers and it can indeed be used for communicative efficiency. However, it appears that as sign languages evolve they find ways to devise linguistic tools that allow for more simultaneous and iconic linguistic structures (e.g., constructed action, depicting signs, lexical signs, pointing for cohesive structures, etc.) to improve the efficiency of a communicative system. In line with previous research, we have gained evidence that while linguistic communication in the visual modality might start as holistic representation through imagistic iconicity and move toward segmentability in a linear manner as it first emerges, later linguistic systems will recruit diagrammatic iconicity for combining meaning elements in simultaneous constructions. We contribute to existing research by providing first insights on the emergent trajectory of a modality-specific property, such as simultaneity through the use of multiple articulators for complex event encoding. Our findings suggest that sign languages constitute linguistic systems that have evolved to take advantage of the affordances of the visual modality, such as the use of multiple articulators for distinct meaning encoding and building devices such as diagrammatic iconicity for efficient communication.

4.7. Acknowledgments

We wish to thank the participants for their participation and authorization to publish their images. We also wish to thank Giulia Demartis and Chiara Madia for data collection and coding. We wish to thank Barbara Pennacchi for drawing the items for the experiment. Furthermore, we wish to thank three anonymous reviewers who have contributed to the paper with their valuable feedback and suggestions during the revision process.

Chapter 5

General discussion and conclusion

The goal of the present thesis was to propose a new perspective on the function of the visual modality-specific properties of simultaneity and iconicity in sign languages. The general premise of my research was that signers would achieve communicative efficiency by making use of simultaneous and iconic constructions and depicting strategies to encode information, rather than relying solely on the linear organization of information. I pointed out that although the properties of simultaneity and iconicity in sign languages are perfectly suited for clustering related meanings closer together not only in a linear but also in a simultaneous fashion, no systematic inquiry into whether they are used to do so in order to achieve communicative efficiency had been undertaken. Thus, I argued that studying how communicative efficiency is achieved in sign languages was crucial for a more complete understanding of how linguistic structure adapts to fit communicative efficiency pressures and whether linguistic modality plays a role in how it is achieved. Thus, in the present thesis I set out to explore whether signers of LIS actually used the modality-specific properties of simultaneity and iconicity when constrained by pressures of communicative efficiency. I approached this topic by means of three complementary research questions aimed at assessing how communicative efficiency is achieved in LIS (Italian Sign Language) from three complementary perspectives:

(RQ1) Are visual modality-specific properties in the form of simultaneous and iconic constructions recruited for efficient communication in LIS?

(RQ2) What linguistic strategies are used for efficient communication in LIS?

(RQ3) Do simultaneous and iconic constructions constitute an emergent property of sign languages or a mere affordance of the visual modality?

5.1. Summary of the main findings

The first research question (RQ1) of this thesis was: *Are visual modality-specific properties in the form of simultaneous and iconic constructions recruited for efficient communication in LIS?* To answer this question, in **Chapter 2** I experimentally explored whether simultaneous and iconic constructions were used by LIS signers in a context that pushed signers to respond to the pressures of communicative efficiency by increasing the number of the information units to be expressed in the stimuli. I hypothesized that simultaneous and iconic constructions in sign languages, and in LIS in this specific case, might be used for clustering related meanings closer together, leading to faster representation access and thus efficient communication, similar to the way spoken languages use dependency distance minimization, where related meanings are clustered closer together on a linear scale (Gibson et al., 2019; Hawkins, 2004). If my hypothesis were true, the simultaneous constructions (quantified as the kinematic simultaneity of two or more articulators encoding distinct semantic information units) used by signers would likely increase as a function of increases in the information density of the event requiring encoding. I further assessed whether the information density of simultaneous constructions (quantified as the specific number of distinct semantic information units encoded simultaneously) also increased with the increase in the information load. The results of the study did indeed show that as the amount of information that had to be communicated increased, not only did the length of the encodings increase but so did the use of simultaneous constructions as well as the information density of these constructions.

This study constitutes the first systematic study that assesses the role of simultaneous and iconic constructions in communicative efficiency in sign languages. Furthermore, it also presents for the first time evidence that the visual modality-specific properties of simultaneity and iconicity are taken advantage of in a linguistic system to accommodate pressures to communicate efficiently through the simultaneous encoding of information pertaining to the same event. These findings highlight the role of the linguistic modality in achieving one of the core functions of language, i.e., communicative efficiency.

The second research question (RQ2) addressed was: *What linguistic strategies are used for efficient communication in LIS?* To answer this question, in **Chapter 3** I used the same data as in **Chapter 2** but this time I analyzed the linguistic strategies used by signers with a particular focus on the depicting strategy of constructed action (CA) and its combination with other strategies. I focused on this strategy for two reasons. First, I argued that the imagistic and diagrammatic properties of CA and the possibility of combining it with other linguistic strategies to depict multiple elements of the event(s) in a single iconic representation made CA a plausible linguistic strategy for achieving communicative efficiency. Second, my aim was to assess use of CA in an experimental setting to see if there was evidence that CA can be used as a referential resource outside of narrative discourse and its evaluative function. I hypothesized that if CA were used to achieve communicative efficiency, its use either alone or in combination with other strategies would increase as informative demands increased. The results of the study revealed that CA was used to a great extent in experimental settings, indicating that this strategy is not limited to narrative discourse. Results further showed that the use of CA on its own and in combination with other linguistic strategies (lexical signs, depicting constructions, pointing) increased as the amount of information requiring encoding increased, indicating that signers recruited the rich iconic potential that such a strategy provides to achieve efficient communication. The present findings highlight the role of the depicting strategy of CA, which historically has been mostly marginalized in the study of language, and put it at the center of linguistic expression in light of its capacity for efficient communication.

The third research question (RQ3) was: *Do simultaneous and iconic constructions constitute an emergent property of sign languages or a mere affordance of the visual modality?* To answer this question, in **Chapter 4** I set out to explore whether it was possible to detect quantitative and qualitative differences in terms of how simultaneous and iconic constructions are used by signers, who employ simultaneity and iconicity as part of their linguistic system, and by silent gesturers, whose only option is to use these properties as a general affordance of the visual modality. I hypothesized that if simultaneous constructions constituted a linguistic property reflecting simultaneous compositionality that has evolved for the function of communicative efficiency, signers would be likely to use these constructions more often and in a more complex way than silent gesturers, allowing signers to produce more informationally dense simultaneous constructions. Indeed, the results revealed that this was the case. The simultaneous constructions used by signers occurred more frequently and were informationally denser than those used by silent gesturers.

Qualitative analyses revealed that while signers employed diagrammatic iconicity to combine multiple related meaning elements in simultaneous constructions, silent gesturers were more limited in their encodings due to relying primarily on imagistic iconicity and representing referents holistically. Furthermore, the results showed that signers were also more redundant in their encodings than silent gesturers by virtue of their encodings being more simultaneous and more informationally dense than those of silent gesturers. Taken together, these findings show that the general affordances of the visual modality allow for simultaneous and iconic encoding, but it is considerably more limited when not used as part of a linguistic system. When used within a linguistic system, signers can take advantage of simultaneous compositionality to organize information when dealing with an increasing information load in order to achieve communicative efficiency by means of clustering related meanings closer together and ensuring their cohesion by means of redundancy. These findings are crucial for supporting the case that simultaneous and iconic constructions are a linguistic resource for communicative efficiency because they show that the increase in the use of such constructions in the face of increasing information load was not simply due to general affordances of the visual modality. On a more general scale, the present findings contribute to the existing research on language evolution by shifting attention from the emergence of linearization to the emergence of simultaneity, a line of research we still know very little about.

5.2. General discussion

On the whole, the findings of the present thesis provide evidence that simultaneous and iconic constructions in a sign language such as LIS constitute a linguistic property that has potentially evolved and is recruited for achieving communicative efficiency. Below I discuss the implications of these findings regarding the view of language as an adaptive and semiotically diverse phenomenon. I then discuss the contribution of iconicity and simultaneity to language as a communicatively efficient system by speculating about their potential cognitive advantages for language processing in the visual modality. Finally, I speculate about the evolutionary trajectory of iconic linguistic strategies that allow more flexibility in simultaneous constructions than afforded by the visual modality alone. I conclude this section with a discussion of the limitations of the present thesis and directions for future research.

5.2.1. Towards a better understanding of language as a semiotically diverse phenomenon

With the present thesis, I hope to have contributed further evidence for the need to expand the existing narrow view of language as a linear, categorical and discrete system by demonstrating that simultaneity and iconicity also constitute core linguistic properties by showing that they can be used flexibly to achieve communicative efficiency (**Chapters 2 and 3**) and that they evolve in a linguistic system to accommodate this communicative function (**Chapter 4**). These claims are in line with recent expanded views of language as a semiotically diverse and adaptive multimodal language system, opposing earlier views based on the linguistic properties mostly typical of the modality-specific properties of only speech or text. Below I outline the historical context of the study of language and the implications for the linguistic status of simultaneous and iconic properties.

Up until relatively recently, the study of *language* has been limited to what can be derived from the study of texts or speech at best (Vigliocco et al., 2014). As a result, the simultaneous, gradient and iconic properties of the visual modality were considered para-linguistic, gestural, and/or non-linguistic and accordingly outside of the general domain of *language,* even in sign languages (for more on this issue, see Demey et al., 2008; Dotter, 2018; Fuks, 2014; Kendon, 2001, 2004, 2008, 2014; Müller, 2018). This is not surprising if we take into consideration that at the beginning of sign language research some seventy years ago, it was crucial to first demonstrate that sign languages were fully-fledged languages on par with languages in the spoken modality. As a consequence, the differences between the two language modalities, although acknowledged, were given less prominence more often than not. Accordingly, the modality-specific properties were either overwhelmingly ignored (e.g., constructed action) or they were equated to spoken language phenomena (e.g., depicting constructions considered akin to “classifiers” in spoken languages) (see Vermeerbergen, 2006 for a review).

The scene has been rapidly changing in the last couple of decades, but the divide between linguistic and non-linguistic or gestural properties prevails (e.g., Cormier et al., 2012; Duncan, 2005; Goldin-Meadow & Brentari, 2017; Johnston & Schembri, 1999; Liddell, 2003). Importantly, this distinction often stems from a spoken language biased view with regard to what can and cannot be considered linguistic by acknowledging that only arbitrary, categorical, and linear combinatorial and compositional properties carry linguistic status (for a discussion, see Kendon, 2008). A thought experiment proposed by Vigliocco et al. (2014, p. 3) clearly demonstrates the fallacy of such an assumption: "What if the study of language had started from signed language rather than spoken language?" If this had been the case, it is likely that the iconic, gradient and simultaneous properties of sign languages would have gotten more attention in linguistics and would have been regarded as core properties of sign languages.

Once we put aside the ideology of what *language* is and instead look at language in light of its communicative functions, such as communicative efficiency, it becomes evident that depicting strategies which employ iconicity, gradience and simultaneity may prove to be useful for linguistic encoding (Brennan, 1990, 1992; Cuxac, 1999, 2000, 2003; Fuks, 2014; Russo, 2004). The results reported in **Chapter 3** show for the first time that CA constitutes a general referential device in the visual linguistic system outside of a narrative setting, and as such it should be considered as belonging at the core of linguistic expression and on par with other linguistic strategies available for signaling reference in sign languages. Furthermore, the findings in **Chapter 4** show that a linguistic system can evolve to be optimized for communicative efficiency by taking advantage of the semiotic diversity available in the visual modality. As such, the line between lexical and gestural or so-called non-linguistic properties becomes somewhat blurry and possibly unnecessary if linguistic expression is conceptualized as a multimodal, semiotically diverse and gradient phenomenon (Capirci et al., 2022; Occhino & Wilcox, 2017; Perniss, 2018). The fact that in **Chapter 3** I found that CA is combined with other linguistic strategies (e.g., lexical signs) to achieve communicative efficiency provides additional evidence for this broader view of language.

There are currently two established approaches that are going in the direction of disregarding the dichotomy between language and gesture. One of them is based on the *semiogenetic/semiological model* (Cuxac, 1999, 2000) which holds that both lexical signs (called frozen signs) and highly iconic structures (including transfer of person, which is comparable to CA) do not stand in opposition to each other but instead are employed based on the intent (illustrative versus non-illustrative) of the signer (Cuxac, 1999, 2000; Cuxac & Pizzuto, 2010; Cuxac & Sallandre, 2007; Fusellier-Souza, 2006; Pizzuto & Corazza, 2000; Russo Cardona & Volterra, 2007; Volterra et al., 2022). Thus, in this account both strategies are considered to be equally linguistic. Another approach that has been gaining prominence is the usage-based approach within cognitive linguistics, which posits that language should be seen as composite utterances encompassing the semiotic diversity of multimodal expressions (e.g., Bybee, 2010; Capirci et al., 2022; Cienki, 2016; Croft & Cruse, 2004; Enfield, 2009; Langacker, 1987; Lepic & Occhino, 2018; Wilcox, 2004; Wilcox & Occhino, 2016, among others). Such an approach eliminates the need to divide between language and gesture both in sign and spoken languages. While both of these approaches use different terminology, the very core of both stems from viewing language as an act of uttering that is based on the cognitive capacities of humans.

Taken together, the findings of this thesis support this broader view of language and goes beyond existing evidence by showing that signers evolve and recruit simultaneous and iconic constructions to achieve communicative efficiency. As such, these constructions ought to be recognized as fundamental linguistic properties and studied more thoroughly with regard to their expressive capacity and their role in communication.

5.2.2. Cognitive advantages of iconicity and simultaneity in communicative efficiency

Adopting a broad view of language and focusing on the communicative function of visual modality-specific properties allowed me to gather evidence that supports the notion that simultaneity and iconicity constitute linguistic properties whose function is to achieve communicative efficiency in a sign language. However, a question remains regarding the exact contributions of simultaneous and iconic constructions to language as a communicatively efficient system. In the remainder of this section, I discuss the potential cognitive advantages that iconicity and simultaneity provide, namely easing the processing involved in the production and comprehension of utterances in sign languages.

5.2.2.1. Cognitive advantage of iconicity via analogue representation

In this section, I first discuss the processing advantage of iconicity that has been attested in the literature, focusing on the lexical level. I then speculate that this advantage might also potentially hold in processing higher-level representations in the form of simultaneous constructions. I then turn to an existing notion in the literature—that iconicity (as opposed to arbitrariness) can hinder communicative efficiency—and argue that the findings of the present thesis suggest that this notion ought to be reconsidered.

The processing advantage of iconicity in sign languages has been attested so far in comprehension at the lexical level. Research suggests that the salient iconic properties of lexical signs speed up sign recognition in a picture/sign matching task (Thompson et al., 2009) and appear to activate meaning automatically (Thompson et al., 2010). Furthermore, signers exhibit enhanced visual-spatial skills (Malaia & Wilbur, 2014; Parasnis et al., 1996; Secora & Emmorey, 2020) and enhanced visual imagery ability (Emmorey et al., 1993), indicating that iconicity might be advantageous for processing. Some researchers argue that iconicity in lexical signs facilitates processing, as it provides a faster path between form and meaning by directly mapping the linguistic form to human experience (Perniss & Vigliocco, 2014; Sidhu et al., 2020; Thompson et al., 2012; Vinson et al., 2015). However, Emmorey (2014) notes that iconicity in signs is better viewed as a structural mapping between a linguistic form and a mental representation in which human experience is schematized rather than being a direct link to the experience itself. In this view, the more linguistic forms are structurally aligned with the elements of the schematization, the more iconicity benefits processing (see Fig. 1-2).

In a similar vein, I would like to speculate that iconicity can also be beneficial for processing higher-level structures beyond the lexical level in both views. That is, simultaneous and iconic constructions allow multiple event elements and their relationship to be encoded in a way that more directly represents human experience. As such, events that are simultaneously perceived in the world can be simultaneously perceived in the linguistic form. On the other hand, processing linguistically encoded events (as opposed to individual concepts) requires not only a simple mapping between form and meaning but also a schematized mapping between linguistic forms and meaning elements and how they are related. Diagrammatic iconicity in simultaneous constructions can ensure more structural alignment between the linguistic form at a higher level of representation and schematization. As such, diagrammatic iconicity might provide added value with regard to processing simultaneous constructions at the event level by allowing for representations that better adhere to mental imagery, which has been shown to be used by signers for language processing (Emmorey, 1993, 1996; Emmorey et al., 1993; Emmorey & Kosslyn, 1996). Accordingly, given the attested processing advantage for iconicity on the lexical level, it should be expected that iconicity might also help ease processing of higher-level structures, thus contributing to the communicative efficiency of simultaneous constructions.

However, the role of iconicity in communicative efficiency has been controversial. Namely, iconicity is generally considered to be a grounding mechanism for referential communication, as it provides motivated links to real-world experience (Murgiano et al., 2021; Perniss et al., 2010; Perniss & Vigliocco, 2014). However, it has been suggested that iconicity (at the lexical level) can in fact be communicatively inefficient as the vocabulary expands, i.e., the greater the number of meanings that have to be referred to, the greater the chance of overlap in form–meaning mapping between different lexical items (Gasser, 2004; Little, Eryılmaz, et al., 2017). Thus, while iconicity is generally considered to be effective as a grounding mechanism for referential communication, arbitrary forms are considered to be more communicatively efficient than iconic ones, as they ensure greater discriminability (Monaghan et al., 2014; Perniss et al., 2010; Perniss & Vigliocco, 2014; Sidhu &

Pexman, 2018). For example, Perniss et al. (2010) argue that iconicity is used to help link language to our experience while arbitrariness is needed for an efficient communication system:

> We propose that iconicity is exploited in the service of guaranteeing the link between linguistic form and human experience. The variability in the forms and amount of iconicity across languages indicate different manners in which languages can get the balance right between two basic constraints, namely the need to link language to our experience (which would favor iconicity) and the need to have an efficient communication system (which would favor arbitrariness). (p. 13)

This argument is sound, but it views iconicity too narrowly, as it generalizes the constraints of iconicity regarding vocabulary (i.e., lexical level) to the communication system in general. When we extend our notion of iconicity from being solely imagistic but also diagrammatic and move beyond the lexical level, we see that iconicity can be recruited for communicative efficiency, as shown in **Chapters 2 and 3**. Thus, a more nuanced view of iconicity as well as communicative efficiency is required in order to generalize about the former being a hindrance for the latter in languages.

In this thesis, therefore, I take a radically different stance from previous accounts speaking to the inefficiency of iconicity. Instead, I argue that diagrammatic iconicity in sign languages allows information to be organized on higher levels not simply by means of abstract rules but also by means of rules that are grounded in human experience of how different elements relate to each other in the real world. For instance, spoken and sign languages employ diagrammatic iconicity in the form of the iconicity of a sequence (*veni-vidi-vici*) as it establishes a motivated relationship between experience in the world and linguistic form (Perniss, 2007a). In the same vein, I would like to argue that in a sign language diagrammatic iconicity is employed as a tool for communicative efficiency, since it allows the encoding of a unified iconic representation to consist of multiple event elements, which maps more directly onto perceptually simultaneous events and their schematization, thus potentially providing a processing advantage in comprehension. Finally, the simultaneous compositionality mentioned in **Chapter 4** also allows related meaning units to be processed together. Future research is needed to explore whether these assumptions can be confirmed.

5.2.2.2. Cognitive advantage of simultaneity via spatial processing

In this section. I first describe the role of simultaneity in communicative efficiency with respect to the tradeoff between articulatory and cognitive effort in the visual modality. I then speculate that sign languages are optimized for communicative efficiency through the simultaneous structuring of linguistic information to optimize for spatial processing in production and comprehension.

It has been argued that simultaneity is one of the properties that makes sign languages efficient in communication, as it allows the number of signs that need to appear in a linear sequence to be reduced when encoding a message (Bellugi & Fischer, 1972; Myers et al., 2011). At first glance, producing shorter linear utterances does indeed require less articulatory and cognitive effort than producing longer linear utterances, which take longer to produce and process. At the same time, simultaneity in terms of multiple articulators encoding multiple distinct units of information in a single movement segment can be considered more articulatorily costly than producing a single sign. Thus, it is crucial to highlight that the potential economy in articulatory effort in terms of utterance length comes at a price of higher articulatory effort in the form of using simultaneous constructions. Furthermore, whether simultaneity actually results in reduced utterance length is not totally clear. In **Chapter 2**, I showed that higher use of simultaneity in signers did not necessarily lead to shorter utterances. And in **Chapter 4** I showed that, relative to silent gesturers, signers' utterances were longer and that simultaneous constructions led to more redundancy (i.e., information units recurring in multiple simultaneous movement segments in an utterance, see Figure 4-18) rather than optimization to reduce length. Thus, it is possible that signers opt for the more articulatory costly simultaneous constructions not only to reduce overall utterance length but because such constructions might be beneficial in other respects, such as easing processing during production and comprehension.

Languages in different modalities can devise different strategies for processing the linguistic signal due to the different modality affordances and constraints

(Emmorey, 2016). Emmorey (2016, p. 26) highlights that "the auditory system is generally more adept at temporal processing than the visual system, which is better at spatial processing". Corroborating this claim, research has shown that unlike for the spoken signal, when the visual modality is involved, simultaneity can be more advantageous than linearity (Wilson & Emmorey, 1997). For example, presenting visual material (written words) in a simultaneous fashion leads to better recall than presenting the same material in a serial fashion (Frick, 1985). Furthermore, serial recall has been shown to be harder for signers than for speakers (Bavelier et al., 2006; Boutla et al., 2004; Geraci et al., 2008; Rudner & Rönnberg, 2008; Wilson & Emmorey, 2006). Thus, given that sign languages are better suited to spatial rather than temporal processing, they have been shown to exhibit a preference for spatial rather than serial structuring of information (Emmorey, 2016). Words in spoken languages are articulated by arranging phonemes sequentially and then processed temporally, while sign languages chunk lexical signs by simultaneously arranging sign parameters (handshape, location, motion, orientation), which are processed spatially (Brentari, 1999; Klima & Bellugi, 1979; Stokoe, 1960). On a morphological level, the preference for spatial chunking in sign languages has been shown by the common use of simultaneous affixation in the form of superimposing morphological marking on the base form (e.g., verb agreement, verb aspects or numerical incorporation) relative to spoken languages, which use linear affixation (Aronoff et al., 2005; Bellugi & Newkirk, 1981; Emmorey, 1995, 2016; Newport, 1981; Newport & Meier, 1985; Özyürek et al., 2010; Supalla, 1982).

In **Chapter 2,** I provide evidence that signers maximize reliance on spatial processing in higher level structures by clustering related semantic information units simultaneously as informational demands increase. For example, in Figure 2-8 a simultaneous construction with four core elements from the same event constitutes a higher-level representation (i.e., a movement segment simultaneously encoding *a bear holding a cat and the cat kissing the bear*). If the lower level units (i.e., meaning elements encoded by means of different articulators) are arranged simultaneously rather than sequentially, they can be processed spatially. In sign languages, then, clustering the related elements of an event together via simultaneous constructions (i.e., dependency distance minimization) can be optimized for spatial processing at higher levels of representation. As such, simultaneity might constitute a strategy that aids processing in production and comprehension in order to increase communicative efficiency.

Simultaneity might provide yet another cognitive advantage from a production perspective. That is, it might potentially reduce the cognitive load in sign languages by diminishing signers' reliance on temporal processing by allowing them to offload information from working memory to the signing space (Risko & Gilbert, 2016). For example, when encoding an informationally dense event, signers can construct an utterance so that previously encoded semantic information units are stored externally by maintaining them in the signing space as new information is being encoded and thus leading to greater redundancy, as found in **Chapter 4** (e.g., in Figure 4-18, a signer maintaining the holding action of the referent while proceeding to encode new information units).

While this assumption remains speculative, it can be supported by evidence from multimodal language research showing that use of gestures during task-solving helps free up cognitive resources (e.g., Alibali & Di Russo, 1999; Carlson et al., 2007; Chu & Kita, 2011; Goldin-Meadow et al., 2001; Novack et al., 2014) and that cognitively more demanding tasks result in increased gesture use (Melinger & Kita, 2007). Furthermore, gesturing has been shown to facilitate language production and organization (Graham & Argyle, 1975; Graham & Heywood, 1975; Hostetter et al., 2007; Jenkins et al., 2017; Morsella & Krauss, 2004; Rauscher et al., 1996). These findings indicate that offloading information from working memory by externalizing it in the visual modality constitutes a strategy for reducing cognitive load for the language producer. Similarly, using simultaneous constructions during the encoding of informationally dense descriptions might free up cognitive resources while signing. Naturally, empirical research is needed to gather evidence for this hypothesis.

5.2.3. Iconicity and simultaneity evolve for communicative efficiency in a linguistic system

In this thesis, I showed that a linguistic system is needed to take full advantage of simultaneous and iconic constructions for purposes of communicative efficiency (**Chapter 4**). The question that remains is why the affordances of the visual modality in the form of simultaneity and iconicity cannot also be recruited to the same extent for communicative efficiency without a linguistic system? In this section, I speculate that as sign languages evolve, they invent more iconic tools (i.e., depicting linguistic strategies), allowing simultaneity to be used more and to a far greater extent than would be afforded by the visual modality alone.

Iconicity has been argued to play an important role in language emergence, as it has been shown to constitute the starting point for the emergence of a communication system (Caldwell & Smith, 2012; Cuskley & Kirby, 2013; Fay, Ellison, et al., 2014; Fay et al., 2010, 2013; Garrod et al., 2007; Macuch Silva et al., 2019; Motamedi et al., 2019; Perlman et al., 2015; Roberts et al., 2015; Theisen et al., 2010). This research shows that when interlocutors have no possibility of using a conventional communication system, they resort to iconicity to communicate specific concepts. The mainstream consensus is that as communication systems in the visual modality evolve, iconicity erodes and the linguistic forms become more arbitrary in order to optimize for communicative efficiency by reducing production effort and increasing discriminability (Aronoff et al., 2005; Fay et al., 2010; Frishberg, 1975; Garrod et al., 2007; Gasser, 2004; Özyürek et al., 2015; Senghas et al., 2004, 2013, but for research showing that iconicity is maintained or increases see Little, Perlman, et al., 2017; Micklos, 2017).

With respect to lexical signs, they can indeed take advantage of a drift toward less iconic forms, as doing so might allow for less effort in producing a single sign in terms of the duration, size, and number of articulators employed to convey a single concept. Given that sign parameters would be still articulated simultaneously and potentially retain some iconicity, it would be also beneficial for the spatial processing upon which sign languages primarily rely on.

However, why then do even established sign languages use highly iconic strategies such as depicting constructions and constructed action instead of lexical signs only? The answer might lie in the fact that beyond the lexical level, using solely lexical signs in non-iconic linear fashion would be less communicatively efficient, as it would result in a linguistic system based predominantly on linear compositionality, which would be more cognitively demanding in the visual modality (Emmorey, 2016; Wilson & Emmorey, 1997). Given that sign languages are better at processing spatial information than temporal information, as discussed in a previous section (5.2.2.), they should be inclined to develop linguistic tools that allow them to capitalize on properties beneficial to spatial processing. Thus, I would like to speculate that innovating new ways for recruiting iconicity (i.e., in the form of different depicting strategies and their combinations with lexical signs and pointing through diagrammatic iconicity) for the purpose of allowing simultaneous compositionality might function as one such tool.

Supporting evidence for optimization for spatial rather than linear processing, and thus the increase in use of simultaneous structure in language emergence, has been found with regard to referential devices. For example, Kocab et al. (2015) have shown that earlier cohorts of NSL (Nicaraguan Sign Language) signers prefer using lexical devices for reference (i.e., lexical signs), while later cohorts prefer using spatial devices (e.g., body shift, spatially modulated lexical signs). This shift to spatial reference, it is argued, enables sign languages to recruit more simultaneity, since three-dimensional space and multiple articulators can be recruited. Additionally, Stamp & Sandler (2021) have shown that as ABSL (Al-Sayyid Bedouin Sign Language) and ISL (Israeli Sign Language) evolve, they allow increasingly more articulators to be recruited simultaneously for linguistic purposes, including simultaneous reference to multiple elements. However, while these studies suggest that as sign languages evolve they devise more simultaneous structures, the role of iconicity in such simultaneous constructions is not acknowledged. Stamp & Sandler (2021) state that greater simultaneity of expression is made possible through the greater abstraction of the linguistic system. The authors show that as ABSL and ISL evolve they develop more abstract and complex referential devices such as *body* *segmentation* (comparable to *body partitioning*, Dudis, 2004), in which the articulators can be used to refer to multiple referents and actions. The authors argue that *body segmentation* is abstract because the body does not holistically stand in for the referent but is segmented by means of different articulators to refer to multiple referents. Although I did not code for type of referential device, the data in **Chapter 4** are in line with the findings of Stamp & Sandler (2021) with regard to the evolution of referential devices. Specifically, there does indeed seem to be considerably more constructions produced by signers that reflect *body segmentation* than by silent gesturers, who appear to be mostly limited to holistic use of the body. However, I would argue that such a device, i.e., *body segmentation*, does not indicate only greater abstraction in terms of segmenting the holistic body into independent units (i.e., body articulators) but also more sophisticated use of iconicity as described above. Namely, while the body of the signer in such constructions is not used holistically, it capitalizes not only on having a body segmented into multiple body articulators but also the possibility of depicting information via CA and the imagistic iconicity afforded by it. On top of that, the signer uses diagrammatic iconicity to establish a meaningful relation between the articulators employed that maps onto the relation between the meaning elements the articulators refer to. As such, signers can construct more simultaneous and thus more iconic constructions, as doing so reflects the perceptually simultaneous event more directly through *body segmentation* rather than through holistic use of the body, which is more limiting, as shown in **Chapter 4**. Furthermore, signers have different linguistic strategies at their disposal, which can all be combined with each other in a single construction due to diagrammatic iconicity, resulting in an overall iconic representation of a simultaneous event (**Chapter 3**). Thus, iconicity evolves in a linguistic system to maximize its expressive possibilities.

Taken together, these findings serve as an invitation to a more thorough exploration of iconicity and simultaneity and their function as properties that are not simply an affordance that can be lost or maintained during language evolution but rather are able to evolve and adapt to fit the communicative needs of language users. I hope to have demonstrated that iconicity is not only an affordance that allows language to emerge or one that bootstraps language acquisition but also a property that evolves in a linguistic system and operates beyond the purely lexical level to allow for efficient communication through simultaneity.

5.2.4. Limitations and future directions

Taken as a whole, the findings of the present thesis confirmed my main hypothesis, which is that signers use modality-specific properties in the form of simultaneous and iconic constructions to achieve communicative efficiency in a context of increasing informational demands. However, for now, this conclusion must be limited to the encoding of quite specific events, namely animate referents involved in transitive action, and to a specific language, namely LIS. Accordingly, in order to be able to generalize these findings, future research is needed to assess whether the same results also hold for encoding different types of events and in different languages in the visual and spoken modalities. Furthermore, in order to confirm these findings, research in the area of language comprehension is required to test whether simultaneous and iconic constructions are actually beneficial for addressees. Research in the area of language development would provide an understanding of how these linguistic properties are acquired and used. Finally, research on language evolution could provide a more thorough understanding of how and under what pressures simultaneous and iconic constructions evolve. I discuss these future directions below.

5.2.4.1. Evidence from various types of events

In the present thesis, I focused on how signers of LIS encode events that are perceptually simultaneous and involve animate referents and transitive actions. As such, in order to be able to generalize these findings, future research should explore the use of simultaneous and iconic constructions in a context of communicative efficiency with different types of events. For example, as I described in **Chapter 1**, CA is a depicting strategy used mainly for depicting animate referents and transitive actions. Accordingly, descriptions of static scenes of various levels of information density and events involving intransitive actions would most probably elicit the use of different linguistic strategies (Engberg-Pedersen, 1993; McDonald, 1982; Özyürek & Perniss, 2010; Pizzuto et al., 2006; Zwitserlood, 2003), which might in turn influence whether and how simultaneous constructions can be used. Accordingly, more research is needed in order to understand the role of simultaneous and iconic constructions for encoding different types of information (e.g., transitive vs. intransitive events, motion vs. static scenes).

5.2.4.2. Comprehension of simultaneous and iconic constructions

While I have shown from the production point of view that simultaneous and iconic constructions can be recruited for efficient communication, a question that remains unanswered is whether such a strategy is also beneficial for comprehension. To assess this, an endeavor for future research would be to compare comprehension of events that are encoded by means of simultaneous and iconic constructions with comprehension of events that are encoded by means of more linearized structures. This could be done by means of a controlled comprehension experiment in which encodings of the same event vary based on whether simultaneous constructions are used. Furthermore, the design can be further enriched by also accounting for the varying information density of the simultaneous constructions used. If simultaneous constructions are also more efficient for comprehension, we can expect faster reaction times when matching the respective event to an encoding. This type of design can also be used to manipulate not only the use of simultaneous and iconic constructions but also utterance length. By manipulating not only simultaneity but also the length of the encodings, it would be possible to tease apart whether simultaneous constructions are beneficial in comprehension (if they are) because of the shorter encodings or whether they are beneficial due to the spatial processing involved, as suggested in previous sections of this discussion.

5.2.4.3. Cross-linguistic and multi-modal evidence

Another exciting avenue for future research would be a cross-linguistic comparison of the use of simultaneous and iconic constructions to achieve communicative efficiency in different sign languages. Such an inquiry would provide deeper insights into which strategies are driven by the cognitive and communicative constraints and affordances of the visual modality and which by typological differences between sign languages (Wilkinson, 2009). For example, it has been argued that the preference for dependency distance minimization is particularly strong in spoken languages with fixed word order, e.g., English, relative to languages with more flexible word order, e.g., German (Gildea & Temperley, 2010). However, in spoken languages with flexible word order, like German, other strategies are used to facilitate processing, such as explicit case marking to indicate who did what to whom (Bentz & Christiansen, 2013; Levshina & Moran, 2021; Lupyan & Christiansen, 2002). It is highly likely that the use of simultaneous and iconic constructions will depend on both language modality and the typological features of a given language.

Furthermore, if a multimodal view of language is adopted, then simultaneity constitutes a general property of not only sign languages but of the language faculty in general. Thus, if my *simultaneity for communicative efficiency hypothesis* holds as a general rather than sign language-specific phenomenon, it would be crucial to assess whether users of spoken languages only use linear strategies, i.e., speech alone, to achieve communicative efficiency, or whether they also take advantage of the potential for simultaneity and iconicity in the form of gestures in addition to speech. It is well documented that users of spoken languages employ the body as a resource for expressing meaning—e.g., by using co-speech gestures and facial expressions. The extent to which speakers use the simultaneity afforded by the employment of body articulators in addition to speech to achieve communicative efficiency has not yet been systematically assessed and is ready for exploration. The methodology developed in the present thesis allows for direct cross-linguistic and cross-modal comparisons in the domain of simultaneity and iconicity.

5.2.4.4. Acquisition of simultaneous and iconic constructions

Developmental research indicates that it takes time to acquire the ability to use simultaneous linguistic forms in which multiple elements of the event are encoded simultaneously in a single construction. Specifically, research on motion verbs and complex indicating verbs shows that children use more linearized forms to encode these concepts, while adult signers predominantly use simultaneous forms (Dudis, 2004; Janzen et al., 2001; Meier, 1987; Morgan et al., 2002; Newport, 1981, 1988;

Supalla, 1982). There has been no research on the acquisition of simultaneous constructions of varying information density and whether the acquisition of such simultaneous constructions proceeds comparably to the acquisition of simultaneous linguistic forms for motion verbs and complex indicating verbs. However, some research indicates that combining CA with other linguistic strategies (e.g., lexical signs, DC) is a complex skill that is not yet completely mastered even by the age of 6-7 (Cormier, Smith, & Sevcikova, 2013). Thus, it should be expected that children will produce fewer simultaneous constructions in their encodings of informationally dense events relative to adult signers. An exciting avenue for future research lies in exploring the developmental trajectory of the acquisition of simultaneous constructions of various levels of information density and their strategic use for achieving communicative efficiency. The methodology of the present thesis allows these aspects to be assessed, as the design of the experiment was developed to be child-friendly.

5.2.4.5. Evolutionary trajectory of simultaneous and iconic constructions

While in the present thesis I have attempted to tap into the emergence of simultaneous and iconic constructions by comparing signers and silent gesturers, I cannot comment on the exact trajectory of how this linguistic property emerges and evolves. The findings of the present thesis show that the vast possibility for depiction in sign languages has little to do with the maintenance of iconicity and simultaneity from the initial stages of language emergence. Rather it reflects the evolution of modality-specific affordances into linguistic properties, allowing for more expressive potential in a language through iconicity and simultaneity. The question that logically arises is: how do these properties evolve and what pressures influence their change? Experimental research employing the iterated learning paradigm (Kirby et al., 2008) and assessing the role of combined pressures of interaction and transmission in language evolution in the visual modality has focused, so far, on the lexical level (Motamedi et al., 2019). Assessing these pressures beyond the lexical level presents an opportunity to test the generalizability of the claims that linguistic systems become less iconic and more linearized as they evolve. In turn, assessing the modality-specific properties in emerging sign languages can help to unravel how and why iconicity and simultaneity become linguistic properties. More insights into the evolutionary trajectory of simultaneous and iconic constructions and the pressures that trigger their evolution may corroborate or refute the notion that these properties evolve to be optimized for communicative efficiency, which I have argued for in this thesis. Thus, future research in the areas of emerging sign languages as well as experimental approaches to the study of evolving sign languages in the lab (e.g., iterated learning paradigm) are required.

5.2.4.6. The interplay between simultaneity and linearization

In the present thesis, I provided the first systematic evidence of signers recruiting simultaneous and iconic constructions for communicative efficiency. While I have provided insights in regard to the use of such constructions, the linguistic strategies they can be constructed of, and their emergent nature in the linguistic structure, it remains to be seen how they interact with linear structure in sign languages. That is to say, linearization remains one of the fundamental properties of all languages, including sign languages (Lepic, 2019; Wilkinson, 2016). As shown in the present thesis, even simultaneous constructions are embedded in a linear structure in the linguistic signal. Thus, future research is needed to gain more understanding about how linearization and simultaneity interact in efficient communication and how they are used to shape the information structure of the discourse.

Furthermore, I argued that in sign languages simultaneity can be used to cluster related meanings as close together as possible, a principle known in spoken languages as dependency distance minimization. In spoken languages, this principle is tightly linked to the exclusively linearized structure of the linguistic signal. As such, dependency distances are computed by counting the intervening words between two dependent words. In sign languages, dependency distances cannot be computed in such a straightforward way, as it is necessary to account for not only the linear but also the simultaneous structure of the linguistic signal. An endeavor for future research would be to explore the ways in which dependency distances can be computed for sign languages, while accounting for their modality-specific properties, such as simultaneity. It would be important to assess whether the linear structure used in utterances containing simultaneous constructions also obeys dependency constraints. An innovative approach to dependency distance quantification in sign languages could provide more evidence for the generalizability of the dependency distance minimizations principle across linguistic modalities and provide insights regarding the interplay between linearization and simultaneity.

5.3. Conclusion

In this thesis, I have attempted to bring together different lines of research on sign languages, communicative efficiency, and language evolution to promote our understanding of the function that visual modality-specific properties play in communication. I showed that simultaneity and iconicity are recruited in LIS for one of the core functions of language—achieving communicative efficiency. I further showed that use of these properties as part of a linguistic system allows more flexibility and efficiency in expression than afforded by the visual modality alone. As such, this thesis contributes to the field of communicative efficiency by providing, for the first time, novel insights with regard to how it is achieved in LIS, a language that exclusively uses the visual modality for linguistic expression. The findings of the present thesis highlight the fundamental role of depiction, and CA in particular, in linguistic organization in a sign language and its referential capacity. This thesis also contributes to the research on language evolution by shifting attention from the linear to simultaneous emergent properties of linguistic structure. Overall, the findings of the present thesis highlight the notion that iconicity and simultaneity constitute linguistic properties that have evolved in sign languages and are used for communicative efficiency. With this thesis I hope to have provided a new avenue for future research that will explore communicative efficiency through a lens that takes a broader view of the language faculty in order to account for its full expressive capacity and semiotic diversity.

References

Alibali, M. W., & Di Russo, A. A. (1999). The function of gesture in learning to count: More than keeping track. *Cognitive Development*, *14*(1), 37–56. https://doi.org/10.1016/S0885-2014(99)80017-3

Allen, R. J., Baddeley, A. D., & Hitch, G. J. (2006). Is the binding of visual features in working memory resource-demanding? *Journal of Experimental Psychology: General*, *135*(2), 298–313. https://doi.org/10.1037/0096-3445.135.2.298

Aronoff, M., Meir, I., Padden, C. A., & Sandler, W. (2008). The roots of linguistic organization in a new language. *Interaction Studies*, *9*(1), 133–153. https://doi.org/10.1075/is.9.1.10aro

Aronoff, M., Meir, I., & Sandler, W. (2005). The paradox of sign language morphology. *Language*, *81*(2), 301–344.

Aylett, M., & Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech*, *47*(1), 31–56. https://doi.org/10.1177/00238309040470010201

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bartek, B., Lewis, R. L., Vasishth, S., & Smith, M. R. (2011). In search of on-line locality effects in sentence comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(5), 1178–1198. https://doi.org/10.1037/a0024194

Bates, D., Mächler, M., Bolker, B., Walker, S., Christensen, R. H., & Singmann, H. (2015). *lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1–7. 2014*.

Bavelier, D., Newport, E. L., Hall, M. L., Supalla, T., & Boutla, M. (2006). Persistent difference in short-term memory span between sign and speech:

Implications for cross-linguistic comparisons. *Psychological Science*, *17*(12), 1090–1092. https://doi.org/10.1111/j.1467-9280.2006.01831.x

Beckner, C., Pierrehumbert, J. B., & Hay, J. (2017). The emergence of linguistic structure in an online iterated learning task. *Journal of Language Evolution*, *2*(2), 160–176. https://doi.org/10.1093/jole/lzx001

Begby, E. (2017). Language from the ground up: A study of homesign communication. *Erkenntnis*, *82*(3), 693–714. https://doi.org/10.1007/s10670-016-9839-1

Bellugi, U., & Fischer, S. (1972). A comparison of sign language and spoken language. *Cognition*, *1*(2–3), 173–200. https://doi.org/10.1016/0010-0277(72)90018-2

Bellugi, U., & Newkirk, D. (1981). Formal devices for creating new signs in American Sign Language. *Sign Language Studies*, *30*(1), 1–35. https://doi.org/10.1353/sls.1981.0001

Bentz, C., & Christiansen, M. H. (2013). Linguistic adaptation: The trade-off between case marking and fixed word orders in Germanic and Romance languages. In G. Peng & F. Shi (Eds.)*, Eastward flows the great river: Festschrift in honor of Prof. William S-Y. Wang on his 80th birthday* (pp. 45–61). Hong Kong: City University of Hong Kong Press.

Boutla, M., Supalla, T., Newport, E. L., & Bavelier, D. (2004). Short-term memory span: Insights from sign language. *Nature Neuroscience*, *7*(9), 997–1002. https://doi.org/10.1038/nn1298

Boyes-Braem, P. (1981). *Significant features of the handshape in American sign language* [Doctoral dissertation, University of California, Berkeley]. https://www.proquest.com/docview/303125626?pq-origsite=gscholar&fromopenview=true

Boyes-Braem, P., Pizzuto, E., & Volterra, V. (2002). The interpretation of signs by (hearing and deaf) members of different cultures. In R. Schulmeister & H. Reinitzer (Eds.), *Progress in sign language research/Fortschritte in der gebärdensprachforschung* (pp. 187–220). Hamburg: Signum Verlag.

Boyes-Braem, P., & Sutton-Spence, R. (2001). *The hands are the head of the mouth: The mouth as articulator in sign languages*. Gallaudet University Press.

Branchini C. (2014). *On relativization and clefting: An analysis of Italian Sign Language*. Berlin: De Gruyter Mouton. https://doi.org/10.1515/9781501500008

Brennan, M. (1990). *Word formation in BSL* [Doctoral dissertation, Stockholm University]. http://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-63586

Brennan, M. (1992). The visual world of BSL: An introduction by Mary Brennan. *Dictionary of British Sign Language/English.* London: Faber and Faber/BDA.

Brentari, D. (1999). *A Prosodic model of sign language phonology*. MIT Press.

Bybee, J. (2010). *Language, usage and cognition*. Cambridge University Press.

Bybee, J. L. (1985). *Morphology: A study of the relation between meaning and form*. Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.9

Bybee, J. L. (2013). Usage-based theory and exemplar representations of constructions. In T. Hoffman & G. Trousdale (Eds.), *The Oxford handbook of construction grammar* (pp. 49-69).Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195396683.013.0004

Caldwell, C. A., & Smith, K. (2012). Cultural evolution and perpetuation of arbitrary communicative conventions in experimental microsocieties. *PLOS ONE*, *7*(8), e43807. https://doi.org/10.1371/journal.pone.0043807

Campisi, E., & Özyürek, A. (2013). Iconicity as a communicative strategy: Recipient design in multimodal demonstrations for adults and children. *Journal of Pragmatics*, *47*(1), 14–27. https://doi.org/10.1016/j.pragma.2012.12.007

Capirci, O., Bonsignori, C., & Di Renzo, A. (2022). Signed languages: A triangular semiotic dimension. *Frontiers in Psychology*, *12:802911*. https://doi.org/10.3389/fpsyg.2021.802911

Capirci, O., Cattani, A., Rossini, P., & Volterra, V. (1998). Teaching sign language to hearing children as a possible factor in cognitive enhancement. *Journal of Deaf Studies and Deaf Education*, *3*(2), 135–142.

Carlson, R. A., Avraamides, M. N., Cary, M., & Strasberg, S. (2007). What do the hands externalize in simple arithmetic? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(4), 747–756. https://doi.org/10.1037/0278-7393.33.4.747

Carrigan, E. (2016). An experimental investigation of the factors supporting the emergence of spatial agreement in Nicaraguan Sign Language. [Doctoral dissertation, University of Connecticut] . https://opencommons.uconn.edu/dissertations/1299

Cartmill, E. A., Beilock, S., & Goldin-Meadow, S. (2012). A word in the hand: Action, gesture and mental representation in humans and non-human primates. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1585), 129–143. https://doi.org/10.1098/rstb.2011.0162

Casey, S. K. (2003). *“Agreement” in gestures and signed languages: The use of directionality to indicate referents involved in actions* [Doctoral dissertation, University of California, San Diego]. https://www.proquest.com/docview/305338656/abstract/D565BDAC93534A09PQ/1

Chen, B., Ning, A., Bi, H., & Dunlap, S. (2008). Chinese subject-relative clauses are more difficult to process than the object-relative clauses. *Acta Psychologica*, *129*(1), 61–65. https://doi.org/10.1016/j.actpsy.2008.04.005

Christiansen, M. H., & Chater, N. (2016). The Now-or-Never bottleneck: A fundamental constraint on language. *The Behavioral and Brain Sciences*, *39*, e62. https://doi.org/10.1017/S0140525X1500031X

Chu, M., & Kita, S. (2011). The nature of gestures’ beneficial role in spatial problem solving. *Journal of Experimental Psychology: General*, *140*(1), 102–116. https://doi.org/10.1037/a0021790

Cienki, A. (2016). Cognitive linguistics, gesture studies, and multimodal communication. *Cognitive Linguistics*, *27*(4), 603–618. https://doi.org/10.1515/cog-2016-0063

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

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

Clark, H. H. (2019). Depicting in communication. In P. Hagoort (Ed.), *Human language: From genes and brains to behavior* (pp. 235–247). MIT Press.

Clark, H. H., & Gerrig, R. J. (1990). Quotations as demonstrations. *Language*, *66*(4), 764–805. https://doi.org/10.2307/414729

Cook, S. W., Yip, T. K., & Goldin-Meadow, S. (2012). Gestures, but not meaningless movements, lighten working memory load when explaining math. *Language and Cognitive Processes*, *27*(4), 594–610. https://doi.org/10.1080/01690965.2011.567074

Coppola, M., & So, W. C. (2006). The seeds of spatial grammar: Spatial modulation and coreference in homesigning and hearing adults. In D. Bamman, T. Magnitskaia, & C. Zaller (Eds.), *Proceedings of the Boston University Conference on Language Development* (Vol. 30, pp. 119–130). Boston: Cascadilla Press.

Cormier, K., Fenlon, J., & Schembri, A. (2015). Indicating verbs in British Sign Language favour motivated use of space. *Open Linguistics*, *1*(1), 684–70. https://doi.org/10.1515/opli-2015-0025

Cormier, K., Quinto-Pozos, D., Sevcikova, Z., & Schembri, A. (2012). Lexicalisation and de-lexicalisation processes in sign languages: Comparing depicting constructions and viewpoint gestures. *Language & Communication*, *32*(4), 329–348. https://doi.org/10.1016/j.langcom.2012.09.004

Cormier, K., Smith, S., & Sevcikova, Z. (2013). Predicate structures, gesture, and simultaneity in the representation of action in British Sign Language: Evidence from deaf children and adults. *The Journal of Deaf Studies and Deaf Education*, *18*(3), 370–390. https://doi.org/10.1093/deafed/ent020

Cormier, K., Smith, S., & Sevcikova-Sehyr, Z. (2015). Rethinking constructed action. *Sign Language & Linguistics*, *18*(2), 167–204. https://doi.org/10.1075/sll.18.2.01cor

Cormier, K., Smith, S., & Zwets, M. (2013). Framing constructed action in British Sign Language narratives. *Journal of Pragmatics*, *55*, 119–139. https://doi.org/10.1016/j.pragma.2013.06.002

Cortazzi, M. (2014). *Narrative analysis*. Routledge. https://doi.org/10.4324/9781315067421

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*(1), 87–114. https://doi.org/10.1017/S0140525X01003922

Cowan, N. (2010). The magical mystery four: How is working memory capacity limited, and why? *Current Directions in Psychological Science*, *19*(1), 51–57. https://doi.org/10.1177/0963721409359277

Crasborn, O. A., Kooij, E. van der, Waters, D., Woll, B., & Mesch, J. (2008). Frequency distribution and spreading behavior of different types of mouth actions in three sign languages. *Sign Language & Linguistics*, *11*(1), 45–67. https://doi.org/10.1075/sll.11.1.04cra

Croft, P. of L. W., Croft, W., Cruse, D. A., & Cruse, S. L. in L. D. A. (2004). *Cognitive linguistics*. Cambridge University Press.

Cuskley, C., & Kirby, S. (2013). Synesthesia, cross-modality, and language evolution. In *The Oxford handbook of synesthesia* (pp. 869–899). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199603329.001.0001

Cuxac, C. (1999). French Sign Language: Proposition of a structural explanation by iconicity. *Proceedings of the International Gesture Workshop on Gesture-Based Communication in Human-Computer Interaction*, 165–184.

Cuxac, C. (2000). *La langue des signes française (LSF): Les voies de l'iconocité*. Ophrys.

Cuxac, C. (2001). Les langues des signes: Analyseurs de la faculté de langage. *Acquisition et interaction en langue étrangère*, *15*, 11–36. https://doi.org/10.4000/aile.536

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

Cuxac, C., & Pizzuto, E. A. (2010). Emergence, norms and variation in signed languages: Towards a conceptual redefinition. *Langage et societe*, *131*(1), 37–53.

Cuxac, C., & Sallandre, M.-A. (2007). Iconicity and arbitrariness in French sign language – highly iconic structures, degenerated iconicity and diagrammatic iconicity. In E. Pizzuto, P. Pietrandrea, & R. Simone (Eds.), *Verbal and signed languages: Comparing structures, concepts and methodologies* (pp. 13–33). De Gruyter Mouton. https://doi.org/10.13140/RG.2.1.4884.8483

Daneman, M., & Green, I. (1986). Individual differences in comprehending and producing words in context. *Journal of Memory and Language*, *25*(1), 1–18. https://doi.org/10.1016/0749-596X(86)90018-5

Demey, E., Van Herreweghe, M., & Vermeerbergen, M. (2008). Iconicity in sign languages. In K. Willems & L. D. Cuypere (Eds.), *Naturalness and iconicity in language* (pp. 189–214). Amsterdam: John Benjamins. https://doi.org/10.1075/ill.7.11dem

Dingemanse, M. (2011). Ideophones and the aesthetics of everyday language in a West-African society. *Senses and Society*, *6*(1), 77–85. https://doi.org/10.2752/174589311X12893982233830

Dingemanse, M. (2015). Ideophones and reduplication: Depiction, description, and the interpretation of repeated talk in discourse. *Studies in Language. International Journal Sponsored by the Foundation "Foundations of Language"*, *39*(4), 946–970. https://doi.org/10.1075/sl.39.4.05din

Dingemanse, M. (2018). Redrawing the margins of language: Lessons from research on ideophones. *Glossa: A Journal of General Linguistics*, *3*(1):4. 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. https://doi.org/10.1016/j.tics.2015.07.013

Dotter, F. (2018). Most characteristic elements of sign language texts are intricate mixtures of linguistic and non-linguistic parts, aren't they? *Colloquium: New Philologies*, 1–62. https://doi.org/10.23963/cnp.2018.3.1.1

Dudis, P. G. (2002). Grounded blend maintenance as a discourse strategy. In C. Lucas (Ed.), *Turn-Taking, Fingerspelling, and Contact in Signed Languages* (pp. 53–72). Gallaudet University Press. https://muse.jhu.edu/chapter/264329

Dudis, P. G. (2004). Body partitioning and real-space blends. *Cognitive Linguistics*, *15*(2), 223–238. https://doi.org/10.1515/cogl.2004.009

Dudis, P. G. (2011). The body in scene depictions. In C. B. Roy (Ed.), *Discourse in Signed Languages* (pp. 3–66). Gallaudet University Press. https://muse.jhu.edu/book/14509

Duncan, S. (2005). Gesture in signing: A case study from Taiwan Sign Language. *Language and Linguistics*, 279–318.

Elliott, L. L. (1962). Backward and forward masking of probe tones of different frequency. *Journal of the Acoustical Society of America*, *34*(8), 1116–1117. https://doi.org/10.1121/1.1918254

Elman, J. L., Bates, E. A., & Johnson, M. H. (1996). *Rethinking innateness: A connectionist perspective on development*. MIT Press.

Emmorey, K. (1993). Processing a dynamic visual-Spatial language: Psycholinguistic studies of American Sign Language. *Journal of Psycholinguistic Research*, *22*(2), 153–187. https://doi.org/10.1007/BF01067829

Emmorey, K. (1995). Processing the dynamic visual-spatial morphology of signed languages. In L. B. Feldman (Ed.), *Morphological Aspects of Language Processing* (pp. 25–54). Psychology Press.

Emmorey, K. (1996). The confluence of space and language in signed languages. In *Language and space* (pp. 171–209). The MIT Press. https://doi.org/10.1146/annurev.anthro.25.1.353

Emmorey, K. (2001). *Language, cognition, and the brain: Insights from sign language research*. Psychology Press. https://doi.org/10.4324/9781410603982

Emmorey, K. (2002). The effects of modality on spatial language: How signers and speakers talk about space. In R. Meier, K. Cormier, & D. Quinto-Pozos (Eds.), *Modality and Structure in Signed and Spoken Languages* (pp. 405-421). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511486777.019

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

Emmorey, K. (2016). Consequences of the Now-or-Never bottleneck for signed versus spoken languages. *The Behavioral and Brain Sciences*, *39*, e70. https://doi.org/10.1017/S0140525X1500076X

Emmorey, K., & Kosslyn, S. M. (1996). Enhanced image generation abilities in deaf signers: A right hemisphere effect. *Brain and Cognition*, *32*(1), 28–44. https://doi.org/10.1006/brcg.1996.0056

Emmorey, K., Kosslyn, S. M., & Bellugi, U. (1993). Visual imagery and visual-spatial language: Enhanced imagery abilities in deaf and hearing ASL signers. *Cognition*, *46*(2), 139–181. https://doi.org/10.1016/0010-0277(93)90017-P

Emmorey, K., & Ozyurek, A. (2014). Language in our hands: Neural underpinnings of sign language and co-speech gesture. In M. S. Gazzaniga, & G. R. Mangun (Eds.), *The cognitive neurosciences* (pp. 657-666). Cambridge, Mass: MIT Press.

Emmorey, K., & Tversky, B. (2002). Spatial perspective choice in ASL. *Sign Language and Linguistics (Online)*, *5*(1), 3–26. https://doi.org/10.1075/sll.5.1.03emm

Enfield, N. J. (2009). *The anatomy of meaning: Speech, gesture, and composite utterances*. Cambridge University Press.

Engberg-Pedersen, E. (1993). *Space in Danish Sign Language: The semantics and morphosyntax of the use of space in a visual language*. Hamburg: Signum.

Engberg-Pedersen, E. (2003). From pointing to reference and predication: Pointing signs, eyegaze, and head and body orientation in Danish Sign Language. In S.

Kita (Ed.), *Pointing: Where language, culture, and cognition meet* (pp. 269–292). Psychology Press.

Fang, G., Gao, W., & Zhao, D. (2007). Large-vocabulary continuous sign language recognition based on transition-movement models. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, *37*(1), 1–9. https://doi.org/10.1109/TSMCA.2006.886347

Fauconnier, G., and Turner, M. B. (1996) Blending as a central process of grammar: Expanded version. In A. Goldberg (Ed.), Conceptual structure, discourse, and language (pp. 113-130). Stanford: Center for the Study of Language and Information. https://ssrn.com/abstract=1293869

Fay, N., Arbib, M., & Garrod, S. (2013). How to bootstrap a human communication system. *Cognitive Science*, *37*(7), 1356–1367. https://doi.org/10.1111/cogs.12048

Fay, N., Ellison, M., & Garrod, S. (2014). Iconicity: From sign to system in human communication and language. *Pragmatics & Cognition*, *22*(2), 244–263. https://doi.org/10.1075/pc.22.2.05fay

Fay, N., Garrod, S., Roberts, L., & Swoboda, N. (2010). The interactive evolution of human communication systems. *Cognitive Science*, *34*(3), 351–386. https://doi.org/10.1111/j.1551-6709.2009.01090.x

Fay, N., Lister, C. J., Ellison, T. M., & Goldin-Meadow, S. (2014). Creating a communication system from scratch: Gesture beats vocalization hands down. *Frontiers in Psychology*, *5*. https://www.frontiersin.org/article/10.3389/fpsyg.2014.00354

Fedorenko, E., Woodbury, R., & Gibson, E. (2013). Direct evidence of memory retrieval as a source of difficulty in non-local dependencies in language. *Cognitive Science*, *37*(2), 378–394. https://doi.org/10.1111/cogs.12021

Fedzechkina, M., Jaeger, T. F., & Newport, E. L. (2012). Language learners restructure their input to facilitate efficient communication. *Proceedings of the National Academy of Sciences*, *109*(44), 17897–17902. https://doi.org/10.1073/pnas.1215776109

Fedzechkina, M., Newport, E. L., & Jaeger, T. F. (2017). Balancing Effort and Information Transmission During Language Acquisition: Evidence From Word Order and Case Marking. *Cognitive Science*, *41*(2), 416–446. https://doi.org/10.1111/cogs.12346

Ferrara, L. (2012). *The grammar of depiction: Exploring gesture and language in Australian Sign Language (Auslan)* [Doctoral dissertation, NTNU]. https://ntnuopen.ntnu.no/ntnu-xmlui/handle/11250/218845

Ferrara, L., & Halvorsen, R. P. (2017). Depicting and describing meanings with iconic signs in Norwegian Sign Language. *Gesture*, *16*(3), 371–395. https://doi.org/10.1075/gest.00001.fer

Ferrara, L., & Hodge, G. (2018). Language as description, indication, and depiction. *Frontiers in Psychology*, *9*. https://www.frontiersin.org/article/10.3389/fpsyg.2018.00716

Ferrara, L., & Johnston, T. (2014). Elaborating who's what: A study of constructed action and clause structure in Auslan (Australian Sign Language). *Australian Journal of Linguistics*, *34*(2), 193–215. https://doi.org/10.1080/07268602.2014.887405

Ferreira, V. S., & Dell, G. S. (2000). Effect of ambiguity and lexical availability on syntactic and lexical production. *Cognitive Psychology*, *40*(4), 296–340. https://doi.org/10.1006/cogp.1999.0730

Ferrer i Cancho, R. (2004). Euclidean distance between syntactically linked words. *Physical Review E*, *70*(5), 056135. https://doi.org/10.1103/PhysRevE.70.056135

Ferrer-i-Cancho, R., & Liu, H. (2014). The risks of mixing dependency lengths from sequences of different length. *Glottotheory*, *5*(2), 143–155. https://doi.org/10.1515/glot-2014-0014

Fischer, S. D., & Siple, P. (Eds.). (1991). Manually coded English: The modality question in signed language development. In *Theoretical Issues in Sign Language Research, Volume 2: Psychology* (pp. 85–109). University of Chicago Press.

Fischer, S., & Janis, W. (1990). Verb sandwiches in American Sign Language. *Current trends in European sign language research, 9*, 279–293.

Fontana, S. (2008). Mouth actions as gesture in sign language. *Gesture*, *8*(1), 104–123. https://doi.org/10.1075/gest.8.1.08fon

Frederiksen, A. T., & Mayberry, R. I. (2016). Who's on first? Investigating the referential hierarchy in simple native ASL narratives. *Lingua*, *180*, 49–68. https://doi.org/10.1016/j.lingua.2016.03.007

Frick, R. W. (1985). Testing visual short-term memory: Simultaneous versus sequential presentations. *Memory & Cognition*, *13*(4), 346–356. https://doi.org/10.3758/BF03202502

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

Fuks, O. (2014). The (non-)random distribution of formational parameters in the established lexicon of Israeli Sign Language (ISL). *Semiotica*, *2014*(199), 125–157. https://doi.org/10.1515/sem-2013-0122

Fusellier-Souza, I. (2006). Emergence and development of signed languages: From a semiogenetic point of view. *Sign Language Studies*, *7*(1), 30–56.

Futrell, R., Levy, R. P., & Gibson, E. (2020). Dependency locality as an explanatory principle for word order. *Language*, *96*(2), 371–412. https://doi.org/10.1353/lan.2020.0024

Futrell, R., Mahowald, K., & Gibson, E. (2015). Large-scale evidence of dependency length minimization in 37 languages. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(33), 10336–10341. https://doi.org/10.1073/pnas.1502134112

Garcia, B., & Sallandre, M.-A. (2020). Contribution of the semiological approach to deixis–anaphora in sign language: The key role of eye-gaze. *Frontiers in Psychology*, *11*: 583763 https://www.frontiersin.org/article/10.3389/fpsyg.2020.583763

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. https://doi.org/10.1080/03640210701703659

Gasser, M. (2004). The origins of arbitrariness in language. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *26*(26). https://escholarship.org/uc/item/34g8355v

Geraci, C. (2015). Italian Sign Language. In J. B. Jepsen, G. De Clerck, S. Lutalo-Kiingi, & W. B. McGregor (Eds.), *Sign Languages of the World* (pp. 473-510). De Gruyter. https://doi.org/10.1515/9781614518174-025

Geraci, C., Gozzi, M., Papagno, C., & Cecchetto, C. (2008). How grammar can cope with limited short-term memory: Simultaneity and seriality in sign languages. *Cognition*, *106*(2), 780–804. https://doi.org/10.1016/j.cognition.2007.04.014

Gernsbacher, M. A. (1997). Coherence cues mapping during comprehension. In J.Costermans & M. Fayol (Eds.), *Processing interclausal relationships: Studies in the production and comprehension of text* (pp. 3–21). Lawrence Erlbaum Associates, Inc.

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, *68*(1), 1–76. https://doi.org/10.1016/S0010-0277(98)00034-1

Gibson, E. (2000). The dependency locality theory: A distance-based theory of linguistic complexity. In *Image, language, brain: Papers from the first mind articulation project symposium* (pp. 94–126). The MIT Press.

Gibson, E., Futrell, R., Piantadosi, S. P., Dautriche, I., Mahowald, K., Bergen, L., & Levy, R. (2019). How efficiency shapes human language. *Trends in Cognitive Sciences*, *23*(5), 389–407. https://doi.org/10.1016/j.tics.2019.02.003

Gibson, E., Piantadosi, S. T., Brink, K., Bergen, L., Lim, E., & Saxe, R. (2013). A Noisy-Channel Account of Crosslinguistic Word-Order Variation. *Psychological Science*, *24*(7), 1079–1088. https://doi.org/10.1177/0956797612463705

Gibson, E., & Wu, H.-H. I. (2013). Processing Chinese relative clauses in context. *Language and Cognitive Processes*, *28*(1–2), 125–155. https://doi.org/10.1080/01690965.2010.536656

Gildea, D., & Temperley, D. (2010). Do grammars minimize dependency length? *Cognitive Science*, *34*(2), 286–310. https://doi.org/10.1111/j.1551-6709.2009.01073.x

Goldin-Meadow, S. (2015). The impact of time on predicate forms in the manual modality: Signers, homesigners, and silent gesturers. *Topics in Cognitive Science*, *7*(1), 169–184. https://doi.org/10.1111/tops.12119

Goldin-Meadow, S., & Beilock, S. L. (2010). Action's influence on thought: The case of gesture. *Perspectives on Psychological Science*, *5*(6), 664–674. https://doi.org/10.1177/1745691610388764

Goldin-Meadow, S., & Brentari, D. (2017). Gesture, sign, and language: The coming of age of sign language and gesture studies. *Behavioral and Brain Sciences*, *40*, E46. https://doi.org/10.1017/S0140525X15001247

Goldin-Meadow, S., McNeill, D., & Singleton, J. (1996). Silence is liberating: Removing the handcuffs on grammatical expression in the manual modality. *Psychological Review*, *103*(1), 34-55. https://doi.org/10.1037/0033-295X.103.1.34

Goldin-Meadow, S., Nusbaum, H., Kelly, S. D., & Wagner, S. (2001). Explaining math: Gesturing lightens the load. *Psychological Science*, *12*(6), 516–522. https://doi.org/10.1111/1467-9280.00395

Goldin-Meadow, S., So, W. C., Özyürek, A., & Mylander, C. (2008). The natural order of events: How speakers of different languages represent events nonverbally. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(27), 9163–9168. https://doi.org/10.1073/pnas.0710060105

Graham, J. A., & Argyle, M. (1975). A cross-cultural study of the communication of extra-verbal meaning by gestures. *International Journal of Psychology*, *10*(1), 57–67. https://doi.org/10.1080/00207597508247319

Graham, J. A., & Heywood, S. (1975). The effects of elimination of hand gestures and of verbal codability on speech performance. *European Journal of Social Psychology*, *5*(2), 189–195. https://doi.org/10.1002/ejsp.2420050204

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

Grodner, D., & Gibson, E. (2005). Consequences of the serial nature of linguistic input for sentenial complexity. *Cognitive Science*, *29*(2), 261–290. https://doi.org/10.1207/s15516709cog0000_7

Grote, E., & Linz, K. (2003). The influence of sign language iconicity on semantic conceptualization. In W. G. Müller & O. Fischer (Eds.), *From sign to signing* (pp. 23–40). Amsterdam: John Benjamins. https://doi.org/10.1075/ill.3.05gro

Haiman, J. (1980). The iconicity of grammar: Isomorphism and motivation. *Language*, *56*(3), 515–540.

Haiman, J. (1985). *Iconicity in Syntax: Proceedings of a Symposium on Iconicity in Syntax, Stanford, June 24-6, 1983*. John Benjamins Publishing.

Hawkins, J. A. (2004). *Efficiency and Complexity in Grammars*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199252695.001.0001

Hawkins, J. A. (2014). *Cross-Linguistic Variation and Efficiency*. OUP Oxford.

Hockett, C. F. (1978). In Search of Jove's Brow. *American Speech*, *53*(4), 243–313. https://doi.org/10.2307/455140

Hockett, C. F., & Hockett, C. D. (1960). The Origin of Speech. *Scientific American*, *203*(3), 88–97.

Hodge, G., & Ferrara, L. (2014). Showing the story: Enactment as performance in Auslan narratives. *Selected Papers from the 44th Conference of the Australian Linguistic Society, 2013*, *44*, 372–397. http://hdl.handle.net/11343/40973

Hodge, G., Ferrara, L., & Anible, B. D. (2019). The semiotic diversity of doing reference in a deaf signed language. *Journal of Pragmatics*, *143*, 33–53. https://doi.org/10.1016/j.pragma.2019.01.025

Hodge, G., & Johnston, T. (2014). Points, depictions, gestures and enactment: Partly lexical and non-lexical signs as core elements of single clause-like units in Auslan (Australian Sign Language). *Australian Journal of Linguistics*, *34*(2), 262–291. https://doi.org/10.1080/07268602.2014.887408

Holler, J., & Wilkin, K. (2011). An experimental investigation of how addressee feedback affects co-speech gestures accompanying speakers' responses. *Journal of Pragmatics*, *43*(14), 3522–3536. https://doi.org/10.1016/j.pragma.2011.08.002

Holt, E. (2000). Reporting and reacting: Concurrent responses to reported speech. *Research on Language and Social Interaction*, *33*(4), 425–454. https://doi.org/10.1207/S15327973RLSI3304_04

Hostetter, A. B., Alibali, M. W., & Kita, S. (2007). I see it in my hands' eye: Representational gestures reflect conceptual demands. *Language and Cognitive Processes*, *22*(3), 313–336. https://doi.org/10.1080/01690960600632812

Hsia, H. J. (1977). Redundancy: Is it the lost key to better communication? *AV Communication Review*, *25*(1), 63–85.

Hsiao, F., & Gibson, E. (2003). Processing relative clauses in Chinese. *Cognition*, *90*(1), 3–27. https://doi.org/10.1016/S0010-0277(03)00124-0

Irwin, D. E., & Andrews, R. V. (1996). Integration and accumulation of information across saccadic eye movements. *Attention and performance XVI: Information integration in perception and communication*, 125–156.

Jaeger, T.F. (2010). Redundancy and reduction: Speakers manage density. *Cognitive Psychology*, *61*, 23–62.

Jaeger, T. F. (2006). *Redundancy and syntactic reduction in spontaneous speech* [Doctoral dissertation, Stanford University].

Jaeger, T. F., & Tily, H. (2011). On language 'utility': Processing complexity and communicative efficiency. *Wiley Interdisciplinary Reviews. Cognitive Science*, *2*(3), 323–335. https://doi.org/10.1002/wcs.126

Jakobson, R. (1960). Style in language. In T. A. Sebeok (Ed.), *Closing statement: Linguistics and poetics* (pp. 350–377). Cambridge: Massachusetts Institute of Technology Press.

Jantunen, T. (2017). Constructed action, the clause and the nature of syntax in Finnish Sign Language. *Open Linguistics*, *3*(1), 65–85. https://doi.org/10.1515/opli-2017-0004

Janzen, T., O'Dea, B., & Shaffer, B. (2001). The construal of events: Passives in American Sign Language. *Sign Language Studies*, *1*(3), 281–310. https://doi.org/10.1353/sls.2001.0009

Jarque, M. J., & Pascual, E. (2016). Mixed viewpoints in factual and fictive discourse in Catalan Sign Language narratives. In B. Dancygier, W. Lu, & A. Verhagen (Eds.), *Viewpoint and the fabric of meaning: Form and use of viewpoint tools across languages and modalities* (pp. 259–280). De Gruyter Mouton.

Jenkins, T., Coppola, M., & Coelho, C. (2017). Effects of gesture restriction on quality of narrative production. *Gesture, 16*(3), 416-431. https://doi.org/10.1075/gest.00003.jen

Jiang, J., & Liu, H. (2015). The effects of sentence length on dependency distance, dependency direction and the implications-based on a parallel English-Chinese dependency treebank. *Language Sciences*, *50*, 93–104. https://doi.org/10.1016/j.langsci.2015.04.002

Johnston, T., & Schembri, A. C. (1999). On defining lexeme in a signed language. *Sign Language & Linguistics*, *2*(2), 115–185. https://doi.org/10.1075/sll.2.2.03joh

Karadöller, D. Z., Sumer, B., Ünal, E., & Ozyurek, A. (2020). Sign advantage for children: Signing children's spatial expressions are more informative than speaking children's speech and gestures combined. *In the 45th Annual Boston University (Virtual) Conference on Language Development (BUCLD 45).*Boston, MA.

Kegl, J., Senghas, A., & Coppola, M. (1999). Creation through contact: Sign language emergence and sign language change in Nicaragua. In M. DeGraff, (Ed.), *Language creation and language change: Creolization, diachrony, and development* (pp. 179–237). Cambridge, MA: MIT Press.

Kendon, A. (2001). Gesture as communication strategy. *Semiotica*, *135*, 191–209. https://doi.org/10.1515/semi.2001.060

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

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

Kendon, A. (2014). Semiotic diversity in utterance production and the concept of ‘Language’. *Philosophical Transactions of Royal Society B*, *369*(1651), 20130293.

Kimmelman, V., Lint, V. de, Vos, C. de, Oomen, M., Pfau, R., Vink, L., & Aboh, E. O. (2019). Argument structure of classifier predicates: Canonical and non-canonical mappings in four sign languages. *Open Linguistics*, *5*(1), 332–353. https://doi.org/10.1515/opli-2019-0018

Kirby, S. (2000). Syntax without natural selection: How compositionality emerges from vocabulary in a population of learners. In C. Knight, (Ed.), *The Evolutionary emergence of language: Social function and the origins of linguistic form* (pp. 303–323). Cambridge University Press.

Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences*, *105*(31), 10681–10686. https://doi.org/10.1073/pnas.0707835105

Kirby, S., Griffiths, T., & Smith, K. (2014). Iterated learning and the evolution of language. *Current Opinion in Neurobiology*, *28*, 108–114. https://doi.org/10.1016/j.conb.2014.07.014

Kirby, S., Tamariz, M., Cornish, H., & Smith, K. (2015). Compression and communication in the cultural evolution of linguistic structure. *Cognition*, *141*, 87–102. https://doi.org/10.1016/j.cognition.2015.03.016

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

Kocab, A., Pyers, J., & Senghas, A. (2015). Referential shift in Nicaraguan Sign Language: A transition from lexical to spatial devices. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.01540

Kocab, A., Senghas, A., & Snedeker, J. (2016). The emergence of temporal language in Nicaraguan Sign Language. *Cognition*, *156*, 147–163. https://doi.org/10.1016/j.cognition.2016.08.005

Kurz, K. B., Mullaney, K., & Occhino, C. (2019). Constructed action in American Sign Language: A look at second language learners in a second modality. *Languages*, *4*(4), 90. https://doi.org/10.3390/languages4040090

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). 'lmerTest' package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13). https://doi.org/10.18637/jss.v082.i13

Labov, W., & Waletsky, J. (1967). Narrative analysis: Oral accounts of personal experiences. Essays on the verbal and visual arts. In J. Helm (Ed.), *Proceedings of the 1967 annual spring meeting of the American Ethnological Society* (pp. 12–44). Seattle: University of Washington Press.

Langacker, R. W. (1987). *Foundations of cognitive grammar: Theoretical prerequisites*. (Vol.1). Stanford, CA: Stanford University Press.

Russo Cardona, T., & Volterra., V. (2007). *Le lingue dei segni. Storia e semiotica*. Carocci Editore.

Lepic, R. (2019). A usage-based alternative to "lexicalization" in sign language linguistics. *Glossa: a journal of general linguistics*, 4(1): 23. https://doi.org/10.5334/gjgl.840

Lepic, R. (2015). *Motivation in morphology: Lexical patterns in ASL and English* [Doctoral dissertation, University of California, San Diego]. https://www.proquest.com/docview/1732684214/abstract/E59F6270EB42419EPQ/1

Lepic, R., & Occhino, C. (2018). A construction morphology approach to sign language analysis. In G. Booij (Ed.), *The construction of words: Advances in construction morphology* (pp. 141–172). Springer International Publishing. https://doi.org/10.1007/978-3-319-74394-3_6

Levelt, W. J. M. (1980). On-line processing constraints on the properties of signed and spoken language. In U. Bellugi & M. Studdert-Kennedy (Eds.), *Signed and spoken language: Biological constraints on linguistic form* (pp. 141–160). Verlag Chemie. https://pure.mpg.de/pubman/faces/ViewItemOverviewPage.jsp?itemId=item_66987

Levelt, W. J. M. (1981). The speaker's linearization problem. *Philosophical Transactions of the Royal Society of London*, *295*(1077), 305–315.

Levshina, N. (in press). *Communicative efficiency: Language structure and use*. Cambridge University Press.

Levshina, N., & Moran, S. (2021). Efficiency in human languages: Corpus evidence for universal principles. *Linguistics Vanguard*, *7*(s3). https://doi.org/10.1515/lingvan-2020-0081

Levy, R., Fedorenko, E., & Gibson, E. (2013). The syntactic complexity of Russian relative clauses. *Journal of Memory and Language*, *69*(4), 461–495. https://doi.org/10.1016/j.jml.2012.10.005

Levy, R., & Jaeger, T.F. (2007). Speakers optimize information density through syntactic reduction. B. Schlökopf, J. Platt, T. Hoffman (Eds.), *Advances in neural information processing systems (NIPS)*, *19*, (pp. 849-856*).* MIT Press, Cambridge, MA. https://proceedings.neurips.cc/paper/2006/hash/c6a01432c8138d46ba39957a8250e027-Abstract.html

Liddell, S. K. (1980). *American sign language syntax*. Mouton.

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

Liddell, S. K., Vogt-Svendsen, M., & Bergman, B. (2007). A crosslinguistic comparison of buoys. In M. Vermeerbergen, L. Leeson, & O. A. Crasborn (Eds.), *Simultaneity in signed languages: Form and function* (pp. 187-215). Amsterdam & Philadelphia: John Benjamins. https://doi.org/10.1075/cilt.281.09lid

Lillo-Martin, D. (1995). The point of view predicate in American Sign Language. In K. Emmorey & J. S. Reilly (Eds.), *Language, gesture, and space*. Psychology Press.

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

Little, H., Perlman, M., & Eryilmaz, K. (2017). *Repeated interactions can lead to more iconic signals*. 760–765. https://pure.mpg.de/pubman/faces/ViewItemOverviewPage.jsp?itemId=item_2441780

Liu, H. (2008). Dependency distance as a metric of language comprehension difficulty. *Journal of Cognitive Science*, *9*(2), 159–191.

Liu, H., Xu, C., & Liang, J. (2017). Dependency distance: A new perspective on syntactic patterns in natural languages. *Physics of Life Reviews*, *21*, 171–193. https://doi.org/10.1016/j.plrev.2017.03.002

Loew, R. C. (1984). *Roles and reference in American Sign Language: A developmental perspective* [Ph.D., University of Minnesota]. https://www.proquest.com/docview/303294857/abstract/56A9048330F84A1DPQ/1

Louwerse, M. M., & Jeuniaux, P. (2010). The linguistic and embodied nature of conceptual processing. *Cognition*, *114*(1), 96–104. https://doi.org/10.1016/j.cognition.2009.09.002

Lu, Q., Xu, C., & Liu, H. (2016). Can chunking reduce syntactic complexity of natural languages? *Complexity*, *21*, 33–41. https://doi.org/10.1002/cplx.21779

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*(6657), 279–284. https://doi.org/10.1038/36846

Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods*, *49*(4), 1494–1502. https://doi.org/10.3758/s13428-016-0809-y

Lupyan, G., & Christiansen, M. H. (2002). Case, word order, and language learnability: Insights from connectionist modeling. In W. D. Gray & C.D. Schunn (Eds.), *Proceedings of the Twenty-Fourth Annual Conference of the Cognitive Science Society* (pp. 596-601). Routledge.

Macuch Silva, V., Holler, J., Ozyurek, A., & Roberts, S. G. (2019). Multimodality and the origin of a novel communication system in face-to-face interaction.

Royal Society Open Science, *7*(1), 182056. https://doi.org/10.1098/rsos.182056

Mahowald, K., Fedorenko, E., Piantadosi, S. T., & Gibson, E. (2013). Info/information theory: Speakers choose shorter words in predictive contexts. *Cognition*, *126*(2), 313–318. https://doi.org/10.1016/j.cognition.2012.09.010

Malaia, E., & Wilbur, R. B. (2014). Enhancement of spatial processing in sign-language users. In D. R. Montello, K. E. Grossner, & D. G. Janelle (Eds.), *Space in mind: Concepts for spatial learning and education* (pp. 159–172). MIT Press.

Manhardt, F., Özyürek, A., Sümer, B., Mulder, K., Karadöller, D. Z., & Brouwer, S. (2020). Iconicity in spatial language guides visual attention: A comparison between signers' and speakers' eye gaze during message preparation. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *46*(9), 1735–1753. https://doi.org/10.1037/xlm0000843

Mather, S., & Winston, E. (1998). Spatial mapping and involvement in ASL storytelling. In C. Lucas (Ed.), *Pinky extension and eye gaze: Language use in deaf communities* (pp. 183–212). Gallaudet University Press.

McDonald, B. H. (1982). *Aspects of the American Sign Language predicate system* [Doctoral dissertation, State University of New York at Buffalo]. https://www.proquest.com/docview/303276069/abstract/F4FF09D9F90A49DEPQ/1

McFarland, C. E., & Kellas, G. (1974). Mode of input effects on subject-controlled processes. *Journal of Experimental Psychology*, *103*(2), 343–350. https://doi.org/10.1037/h0036819

McNeill, D. (1996). *Hand and mind: What gestures reveal about thought*. University of Chicago Press. https://press.uchicago.edu/ucp/books/book/chicago/H/bo3641188.html

Meier, R. P. (1987). Elicited imitation of verb agreement in American Sign Language: Iconically or morphologically determined? *Journal of Memory* *and Language*, *26*(3), 362–376. https://doi.org/10.1016/0749-596X(87)90119-7

Meir, I., Padden, C. A., Aronoff, M., & Sandler, W. (2007). Body as subject. *Journal of Linguistics*, *43*(3), 531–563. https://doi.org/10.1017/S0022226707004768

Meir, I., Padden, C., Aronoff, M., & Sandler, W. (2013). Competing iconicities in the structure of languages. *Cognitive Linguistics*, *24*(2), 309–343. https://doi.org/10.1515/cog-2013-0010

Melinger, A., & Kita, S. (2007). Conceptualisation load triggers gesture production. *Language and Cognitive Processes*, *22*(4), 473–500. https://doi.org/10.1080/01690960600696916

Metzger, M. (1995). Constructed action and constructed dialogue in American Sign Language. In C. Lucas (Ed.), *Sociolinguistics in Deaf Communities* (pp. 255–271). Gallaudet University Press.

Micklos, A. (2017). Iconic strategies in silent gesture: Perceiving the distinction between nouns and verbs. In *11th international symposium on iconicity in language and literature* (pp. 26–27). Micklos/01132ed58a26165135927d56a1c48204ac63cf55

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81–97. https://doi.org/10.1037/h0043158

Mineiro, A., Carmo, P., Caroça, C., Moita, M., Carvalho, S., Paço, J., & Zaky, A. (2017). Emerging linguistic features of Sao Tome and Principe Sign Language. *Sign Language & Linguistics*, *20*(1), 109–128. https://doi.org/10.1075/sll.20.1.04min

Monaghan, P., Shillcock, R. C., Christiansen, M. H., & Kirby, S. (2014). How arbitrary is language? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1651), 20130299. https://doi.org/10.1098/rstb.2013.0299

Morford, J. P. (2002). The expression of motion events in homesign. *Sign Language & Linguistics*, *5*(1), 55–71. https://doi.org/10.1075/sll.5.1.05mor

Morgan, G. (2002). Children's encoding of simultaneity in British Sign Language narratives. *Sign Language and Linguistics (Online)*, *5*(2), 131–165. https://doi.org/10.1075/sll.5.2.04mor

Morgan, G., Herman, R., & Woll, B. (2002). The development of complex verb constructions in British Sign Language. *Journal of Child Language*, *29*(3), 655–675. https://doi.org/10.1017/s0305000902005184

Morgan, H. E. (2017). *The phonology of Kenyan Sign Language (southwestern dialect)* [Doctoral dissertation, University of California, San Diego]. https://www.proquest.com/docview/2007620741/fulltextPDF/2E1020E4B8A54284PQ/1?accountid=105521

Morgan, H. E., & Mayberry, R. I. (2012). Complexity in two-handed signs in Kenyan Sign Language: Evidence for sublexical structure in a young sign language. *Sign Language & Linguistics, 15*(1), 147-174. https://doi.org/10.1075/sll.15.1.07mor

Morsella, E., & Krauss, R. M. (2004). The role of gestures in spatial working memory and speech. *The American Journal of Psychology*, *117*(3), 411–424. https://doi.org/10.2307/4149008

Motamedi, Y., Schouwstra, M., Smith, K., Culbertson, J., & Kirby, S. (2019). Evolving artificial sign languages in the lab: From improvised gesture to systematic sign. *Cognition*, *192*, 103964. https://doi.org/10.1016/j.cognition.2019.05.001

Müller, C. (2018). Gesture and sign: Cataclysmic break or dynamic relations? *Frontiers in Psychology*, *9*. https://www.frontiersin.org/article/10.3389/fpsyg.2018.01651

Murgiano, M., Motamedi, Y., & Vigliocco, G. (2021). Situating language in the real-world: The role of multimodal iconicity and indexicality. *Journal of Cognition*, *4*(1), 38. https://doi.org/10.5334/joc.113

Myers, J., Tsay, J. S., & Su, S. (2011). Representation efficiency and transmission efficiency in sign and speech. In J. Chang (Ed.), *Language and cognition: Festschrift in honor of James H-Y. Tai on his 70th birthday* (pp. 171–199). Taipei: Crane Publishing.

https://www.semanticscholar.org/paper/Representation-Efficiency-and-Transmission-in-Sign-Myers-Tsay/0fa8691a2178a9f5bc8a5728ca0bdcf4d06f6737

Napoli, D., & Sutton-Spence, R. (2014). Order of the major constituents in sign languages: Implications for all language. *Frontiers in Psychology*, *5*:376. https://doi.org/10.3389/fpsyg.2014.00376

Napoli, D., & Sutton-Spence, R. (2010). Limitations on simultaneity in sign language. *Language*, *86*(3), 647–662. https://doi.org/10.1353/LAN.2010.0018

Newport, E. L. (1981). Constraints on structure: Evidence from American Sign Language. In W.A. Collins, (Ed.), *Aspects of the development of competence*. Psychology Press. https://doi.org/10.4324/9780203780916

Newport, E. L. (1988). Constraints on learning and their role in language acquisition: Studies of the acquisition of American sign language. *Language Sciences*, *10*(1), 147–172. https://doi.org/10.1016/0388-0001(88)90010-1

Newport, E. L., & Meier, R. P. (1985). The acquisition of American Sign Language. In *The crosslinguistic study of language acquisition, Vol. 1: The data; Vol. 2: Theoretical issues* (pp. 881–938). Lawrence Erlbaum Associates, Inc.

Nölle, J., Staib, M., Fusaroli, R., & Tylén, K. (2018). The emergence of systematicity: How environmental and communicative factors shape a novel communication system. *Cognition*, *181*, 93–104. https://doi.org/10.1016/j.cognition.2018.08.014

Novack, M. A., Congdon, E. L., Hemani-Lopez, N., & Goldin-Meadow, S. (2014). From action to abstraction: Using the hands to learn math. *Psychological Science*, *25*(4), 903–910. https://doi.org/10.1177/0956797613518351

Occhino, C. (2017). An introduction to embodied cognitive phonology: Claw-5 handshape distribution in asl and libras. *Complutense Journal of English Studies*, *25*, 69–103.

Occhino, C., & Wilcox, S. (2017). Gesture or sign? A categorization problem. *Behavioral and Brain Sciences*, *40*, E66. https://doi.org/10.1017/S0140525X15003015

Oliveira, F. T. P., & Ivry, R. B. (2008). The representation of action: Insights from bimanual coordination. *Current Directions in Psychological Science*, *17*(2), 130–135. https://doi.org/10.1111/j.1467-8721.2008.00562.x

Ortega, G. (2017). Iconicity and sign lexical acquisition: A review. *Frontiers in Psychology*, *8:* 1280. https://doi.org/10.3389/fpsyg.2017.01280

Ortega, G., & Morgan, G. (2015). Phonological development in hearing learners of a sign language: The influence of phonological parameters, sign complexity, and iconicity. *Language Learning*, *65*(3), 660–688. https://doi.org/10.1111/lang.12123

Ortega, G., & Özyürek, A. (2020). Types of iconicity and combinatorial strategies distinguish semantic categories in silent gesture across cultures. *Language and Cognition*, *12*(1), 84–113. https://doi.org/10.1017/langcog.2019.28

Ortega, G., Sümer, B., & Özyürek, A. (2017). Type of iconicity matters in the vocabulary development of signing children. *Developmental Psychology*, *53*(1), 89–99. https://doi.org/10.1037/dev0000161

Özçalışkan, Ş., Lucero, C., & Goldin-Meadow, S. (2016). Does language shape silent gesture? *Cognition*, *148*, 10–18. https://doi.org/10.1016/j.cognition.2015.12.001

Özyürek, A. (2021). Considering the nature of multimodal language from a crosslinguistic perspective. *Journal of Cognition*, *4*(1), 42. https://doi.org/10.5334/joc.165

Özyürek, A., Furman, R., & Goldin-Meadow, S. (2015). On the way to language: Event segmentation in homesign and gesture. *Journal of Child Language*, *42*(1), 64–94. https://doi.org/10.1017/S0305000913000512

Özyürek, A., & Perniss, P. (2011). Event representations in signed languages. In E. Pederson & J. Bohnemeyer (Eds.), *Event representation in language and Cognition* (pp. 84–107). Cambridge University Press. https://doi.org/10.1017/CBO9780511782039.005

Özyürek, A., & Trabasso, T. (1997). Evaluation during the understanding of narratives. *Discourse Processes*, *23*(3), 305–335. https://doi.org/10.1080/01638539709544995

Özyürek, A., & Woll, B. (2019). Language in the visual modality: Cospeech gesture and sign language. In *Human language: From genes and brain to behavior* (pp. 67–83). MIT Press.

Özyürek, A., Zwitserlood, I., & Perniss, P. (2010). Locative expressions in signed languages: A view from Turkish Sign Language (TD). *Linguistics*, *48*(5), 1111–1145. https://doi.org/10.1515/LING.2010.036

Padden, C. (1988). *Interaction of morphology and syntax in American Sign Language*. Garland.

Padden, C. (1986). Verbs and role-shifting in American Sign Language. *Proceedings of the Fourth National Symposium on Sign Language Research and Teaching*, *44*, 57. https://pages.ucsd.edu/~cpadden/files/Verbs%20and%20Role-Shifting%20in%20ASL.pdf

Padden, C., Meir, I., Hwang, S.-O., Lepic, R., Seegers, S., & Sampson, T. (2013). Patterned iconicity in sign language lexicons. *Gesture*, *13*(3), 287–308. https://doi.org/10.1075/gest.13.3.03pad

Padden, C., Hwang, S.-O., Lepic, R., & Seegers, S. (2015). Tools for language: Patterned iconicity in sign language nouns and verbs. *Topics in Cognitive Science*, *7*(1), 81–94. https://doi.org/10.1111/tops.12121

Parasnis, I., Samar, V., Bettger, J., & Sathe, K. (1996). Does deafness lead to enhancement of visual spatial cognition in children? Negative evidence from deaf nonsigners. *Journal of Deaf Studies and Deaf Education*, *1*(2), 145–152. https://doi.org/10.1093/oxfordjournals.deafed.a014288

Pashler, H. (1988). Familiarity and visual change detection. *Perception & Psychophysics*, *44*(4), 369–378. https://doi.org/10.3758/BF03210419

Pellegrino, F., Coupé, C., & Marsico, E. (2011). Across-language perspective on speech information rate, *Language 87(3),* 539-558. https://doi.org/10.1353/LAN.2011.0057

Penney, C. G. (1989). Modality effects and the structure of short-term verbal memory. *Memory & Cognition*, *17*(4), 398–422. https://doi.org/10.3758/BF03202613

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

Perniss, P. (2007a). *Space and iconicity in German Sign Language (DGS)* [[S.l.] : MPI Series in Psycholinguistics]. https://repository.ubn.ru.nl/handle/2066/30937

Perniss, P. (2007b). Achieving spatial coherence in German Sign Language narratives: The use of classifiers and perspective. *Lingua*, *117*(7), 1315–1338. https://doi.org/10.1016/j.lingua.2005.06.013

Perniss, P. (2018). Why we should study multimodal language. *Frontiers in Psychology*, *9: 1109*. https://www.frontiersin.org/article/10.3389/fpsyg.2018.01109

Perniss, P., & Özyürek, A. (2015). Visible cohesion: A comparison of reference tracking in sign, speech, and co-speech gesture. *Topics in Cognitive Science*, *7*(1), 36–60. https://doi.org/10.1111/tops.12122

Perniss, P., Özyürek, A., & Morgan, G. (2015). The influence of the visual modality on language structure and conventionalization: Insights from sign language and gesture. *Topics in Cognitive Science*, *7*(1), 2–11. https://doi.org/10.1111/tops.12127

Perniss, P., Thompson, R., & Vigliocco, G. (2010). Iconicity as a general property of language: Evidence from spoken and signed languages. *Frontiers in Psychology*, *1*. https://www.frontiersin.org/article/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*(1651), 20130300. https://doi.org/10.1098/rstb.2013.0300

Piantadosi, S. T., Tily, H., & Gibson, E. (2011). Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(9), 3526–3529. https://doi.org/10.1073/pnas.1012551108

Pietrandrea, P., & Russo, T. (2007). Diagrammatic and imagic hypoicons in signed and verbal languages. In E. Pizzuto, P. Pietrandrea, & R. Simone (eds.) (Eds.), *Verbal and signed languages. Comparing structures, constructs and methodologies* (pp. 35–56). Mouton de Gruyter. https://hal.archives-ouvertes.fr/hal-00665341

Pizzuto, E., & Corazza, S. (2000). Segni senza parole: Osservazioni sui 'classificatori' della LIS. In C. Bagnara, G. Chiappini, & M. P. Conte (Eds.), *Viaggio nella città invisibile. Atti del 2º Convegno nazionale sulla lingua italiana dei segni* (pp. 50–59). Edizioni del Cerro.

Pizzuto, E., Rossini, P., Sallandre, M.A., & Wilkinson, E. (2006). Deixis, anaphora and highly iconic structures: Cross-linguistic evidence on American (ASL), French (LSF) and Italian (LIS) Signed Languages. *TISLR9*, 475–495. https://hal.archives-ouvertes.fr/hal-00617907

Porcari Li Destri, G., & Volterra, V. (1995). *Passato e presente. Uno sguardo all'educazione dei sordi in Itali*. Gnocchi Editore, Napoli1995.

Poulin, C., & Miller, C. (1995). On narrative discourse and point of view in quebec sign language. In K. Emmorey & J. Reilly (Eds.), *Language, Gesture, and Space* (pp. 117–131). Psychology Press.

Puupponen, A. (2019). *Understanding nonmanuality – A study on the actions of the head and body in Finnish Sign Language* [Doctoral dissertation, University of Jyväskylä]. https://jyx.jyu.fi/handle/123456789/63646

Quer, J. (2011). Reporting and quoting in signed discourse. In E. Brendel, J. Meibauer, & M. Steinbach (Eds.), *Understanding Quotation:* (pp. 277–302). De Gruyter Mouton https://doi.org/10.1515/9783110240085

Quer, J. (2018). On categorizing types of role shift in Sign languages. *Theoretical Linguistics*, *44*(3–4), 277–282. https://doi.org/10.1515/tl-2018-0020

Quinto-Pozos, D. (2007a). Why does constructed action seem obligatory? An analysis of classifiers and the lack of articulator-referent correspondence. *Sign Language Studies*, *7*(4), 458–506. https://doi.org/10.1353/sls.2007.0027

Quinto-Pozos, D. (2007b). Can constructed action be considered obligatory? *Lingua*, *117*(7), 1285–1314. https://doi.org/10.1016/j.lingua.2005.12.003

Quinto-Pozos, D., & Mehta, S. (2010). Register variation in mimetic gestural complements to signed language. *Journal of Pragmatics*, *42*(3), 557–584. https://doi.org/10.1016/j.pragma.2009.08.004

Race, D. S., & MacDonald, M. C. (2003). The use of "that" in the production and comprehension of object relative clauses. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *25*(25).

Radutzky, E. (1992). *Dizionario bilingue elementare della lingua italiana dei segni. Oltre 2500 significati. Con DVD-ROM*. Kappa.

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. https://doi.org/10.1111/j.1467-9280.1996.tb00364.x

Raviv, L., Meyer, A., & Lev-Ari, S. (2019). Larger communities create more systematic languages. *Proceedings of the Royal Society B: Biological Sciences*, *286*(1907), 20191262. https://doi.org/10.1098/rspb.2019.1262

Remez, R. E., Ferro, D. F., Dubowski, K. R., Meer, J., Broder, R. S., & Davids, M. L. (2010). Is desynchrony tolerance adaptable in the perceptual organization of speech? *Attention, Perception, & Psychophysics*, *72*(8), 2054–2058. https://doi.org/10.3758/BF03196682

Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. *Trends in Cognitive Sciences*, *20*(9), 676–688. https://doi.org/10.1016/j.tics.2016.07.002

Risler, A. (2007). A cognitive linguistic view of simultaneity in process signs in French Sign Language. In M. Vermeerbergen, L. Leeson, & O. Crasborn (Eds.), *Simultaneity in signed languages. Form and function* (pp. 73-101). Amsterdam & Philadelphia: John Benjamins. https://doi.org/10.1075/cilt.281.04ris

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. https://doi.org/10.1016/j.cognition.2015.04.001

Roccaforte, M., & Volterra, V. (2016). 30. La Lingua dei Segni italiana. In *30. La Lingua dei Segni italiana* (pp. 707–728). De Gruyter. https://doi.org/10.1515/9783110360851-032

Rogers, K. L. (2012). *American Sign Language verb categories in constructed action* [Doctoral dissertation, University of Texas Arlington]. https://rc.library.uta.edu/uta-ir/handle/10106/11124

Roy, C. B. (1989). Features of discourse in an American Sign Language lecture. In C. Lucas (Ed.), *The sociolinguistics of the Deaf community* (pp. 231–251). Academic Press. https://doi.org/10.1016/B978-0-12-458045-9.50017-3

Rudner, M., & Rönnberg, J. (2008). Explicit processing demands reveal language modality-specific organization of working memory. *Journal of Deaf Studies and Deaf Education*, *13*(4), 466–484. https://doi.org/10.1093/deafed/enn005

Russo, T. (2004). Iconicity and productivity in sign language discourse: An analysis of three LIS discourse registers. *Sign Language Studies*, *4*(2), 164–197. https://doi.org/10.1353/sls.2004.0006

Sallandre, M., Balvet, A., Besnard, G., & Garcia, B. (2019). Étude exploratoire de la fréquence des catégories linguistiques dans quatre genres discursifs en LSF. *Lidil. Revue de linguistique et de didactique des langues*, *60*, Article 60. https://doi.org/10.4000/lidil.7136

Sandler, W., Aronoff, M., Padden, C., Meir, I., Enfield, N., Kockelman, P., & Sidnell, J. (2014). Language emergence: Al-Sayyid bedouin sign language. *Cambridge Handbook of Linguistic Anthropology*, 250–284.

Sandler, W., & Lillo-Martin, D. (2006). *Sign language and linguistic universals*. Cambridge University Press.

Schembri, A. (2003). Rethinking 'classifiers' in signed languages. In K. Emmorey (Ed.), *Perspectives on classifier constructions in sign languages* (pp. 3–34). Psychology Press.

Schouwstra, M., & de Swart, H. (2014). The semantic origins of word order. *Cognition*, *131*(3), 431–436. https://doi.org/10.1016/j.cognition.2014.03.004

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

Senghas, A. (2019). How language learns: Linking universals to acquisition. *Proceedings of the 43rd Annual Boston University Conference on Language Development*, 1–10.

Senghas, A., & Coppola, M. (2001). Children creating language: How Nicaraguan Sign Language acquired a spatial grammar. *Psychological Science*, *12*(4), 323–328. https://doi.org/10.1111/1467-9280.00359

Senghas, A., Kita, S., & Özyürek, A. (2004). Children creating core properties of language: Evidence from an emerging sign language in Nicaragua. *Science*, *305*(5691), 1779–1782. https://doi.org/10.1126/science.1100199

Senghas, A., & Littman, S. (2004). Segmentation in the expression of motion events in co-speech gesture, Nicaraguan Sign Language (NSL) and Spanish Sign Language (LSE). *Proceedings in Theoretical Issues in Sign Language Research*. TISLR 8, Barcelona, Spain.

Senghas, A., Özyürek, A., & Goldin-Meadow, S. (2010). The evolution of segmentation and sequencing: Evidence from homesign and Nicaraguan Sign Language. In A. D. Smith, M. Schouwstra, B. de Boer, & K. Smith (Eds.), *Proceedings of the 8th International conference on the Evolution of Language* (EVOLANG 8) (pp. 279-289). Singapore: World Scientific. https://doi.org/10.1142/9789814295222_0036

Senghas, A., Özyürek, A., & Goldin-Meadow, S. (2013). Homesign as a way-station between co-speech gesture and sign language: The evolution of segmenting and sequencing. In R. Botha, & M. Everaert (Eds.), *The evolutionary emergence of language: Evidence and inference* (pp. 62-77). Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199654840.003.0004

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, *27*(3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Sidhu, D. M., & Pexman, P. M. (2018). Five mechanisms of sound symbolic association. *Psychonomic Bulletin & Review*, *25*(5), 1619–1643. https://doi.org/10.3758/s13423-017-1361-1

Sidhu, D. M., Vigliocco, G., & Pexman, P. M. (2020). Effects of iconicity in lexical decision. *Language and Cognition: An Interdisciplinary Journal of Language and Cognitive Science*, *12*(1), 164–181. https://doi.org/10.1017/langcog.2019.36

Siple, P., & Fischer, S. D. (Eds.). (1991). *Theoretical issues in sign language research: Psychology* (Vol. 2). University of Chicago Press. https://press.uchicago.edu/ucp/books/book/chicago/T/bo3637114.html

Slonimska, A., Di Renzo, A., & Capirci, O. (2018). *The use of highly iconic structures in Italian Sign Language (LIS) in a narrative context: A developmental perspective* [conference presentation]. International Sign Language Acquisition conference (ICSLA), Istanbul, Turkey.

Slonimska, A., Özyürek, A., & Capirci, O. (2020). The role of iconicity and simultaneity for efficient communication: The case of Italian Sign Language (LIS). *Cognition*, *200*, 104246. https://doi.org/10.1016/j.cognition.2020.104246

Slonimska, A., Özyürek, A., & Capirci, O. (2021). Using depiction for efficient communication in LIS (Italian Sign Language). *Language and Cognition*, *13*(3), 367–396. https://doi.org/10.1017/langcog.2021.7

Smith, K., Kirby, S., & Brighton, H. (2003). Iterated learning: A framework for the emergence of language. *Artificial Life*, *9*(4), 371–386. https://doi.org/10.1162/106454603322694825

Smith, K., Tamariz, M., & Kirby, S. (2013). Linguistic structure is an evolutionary trade-off between simplicity and expressivity. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *35*(35).

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs: General and Applied*, *74*(11), 1–29. https://doi.org/10.1037/h0093759

Stamp, R., & Sandler, W. (2021). The emergence of referential shift devices in three young sign languages. *Lingua*, *257*, 103070. https://doi.org/10.1016/j.lingua.2021.103070

Stokoe, W. (1960). Sign language structure, an outline of the visual communications systems of American deaf. In *Studies in linguistics: Occasional paper* (Vol. 8). Buffalo: Dept. of Anthropology and Linguistics, University of Buffalo.

Studdert-Kennedy, M. (1986). Some developments in research on language behavior. In N. J. Smelser & D. R. Gerstein (Eds.), *Behavioral and social science: Fifty years of discovery* (pp. 208–248). National Academy Press.

Sümer, B. (2015). *Acquisition of spatial language by signing and speaking children: A comparison of Turkish Sign Language (TID) and Turkish* [Radboud University]. https://repository.ubn.ru.nl/handle/2066/140304

Sümer, B., Zwitserlood, I., Perniss, P., & Ozyurek, A. (2013). Acquisition of locative expressions in children learning Turkish Sign Language (TİD) and Turkish. *Current Directions in Turkish Sign Language Research*, 243–272. https://doi.org/10/component/file_2260554/10Sumer-etal-formatted.pdf

Supalla, T. (1986). The classifier system in American sign language. In C. G. Craig (Ed.), *Noun classes and categorization* (pp. 181–214). John Benjamins Publishing Company.

Supalla, T. (2003). Revisiting Visual Analogy in ASL Classifier Predicates. In K. Emmorey (Ed.), *Perspectives on Classifier Constructions in Sign Languages*. Psychology Press.

Supalla, T. (1982). *Structure and acquisition of verbs of motion and location in American sign language* [Doctoral dissertation, University of California, San Diego]. https://www.proquest.com/docview/303234697/BB33A1AD172A4469PQ/1?accountid=105521

Sutton-Spence, R. (2007). Mouthings and simultaneity in British sign language. In M. Vermeerbergen, L. Leeson, & O. Crasborn (Eds.), *Simultaneity in signed languages: Form and function* (pp. 147-162). Amsterdam & Philadelphia: John Benjamins. https://doi.org/10.1075/cilt.281.07sut

Sutton-Spence, R., & Boyes Braem, P. (2013). Comparing the products and the processes of creating sign language poetry and pantomimic improvisations. *Journal of Nonverbal Behavior*, *37*(4), 245–280. https://doi.org/10.1007/s10919-013-0160-2

Tagliamonte, S., Smith, J., & Lawrence, H. (2005). No taming the vernacular! Insights from the relatives in northern Britain. *Language Variation and Change*, *17*(1), 75–112. https://doi.org/10.1017/S0954394505050040

Tamariz, M., & Kirby, S. (2016). The cultural evolution of language. *Current Opinion in Psychology*, *8*, 37–43. https://doi.org/10.1016/j.copsyc.2015.09.003

Tannen, D. (1989). *Talking vcoices: Repetition, dialogue, and imagery in conversational discourse*. Cambridge University Press.

Taub, S. F. (2001). *Language from the body: Iconicity and metaphor in American Sign Language*. Cambridge University Press. https://doi.org/10.1017/CBO9780511509629

Temperley, D. (2007). Minimization of dependency length in written English. *Cognition*, *105*(2), 300–333. https://doi.org/10.1016/j.cognition.2006.09.011

Temperley, D., & Gildea, D. (2018). Minimizing syntactic dependency lengths: Typological/cognitive universal? *Annual Review of Linguistics*, *4*, 67–80. https://doi.org/10.1146/annurev-linguistics-011817-045617

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

Theisen-White, C., Kirby, S., & Oberlander, J. (2011). Integrating the horizontal and vertical cultural transmission of novel communication systems. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *33*. Retrieved from https://escholarship.org/uc/item/40t7578c

Thompson, R. L., Vinson, D. P., & Vigliocco, G. (2009). The link between form and meaning in American Sign Language: Lexical processing effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(2), 550–557. https://doi.org/10.1037/a0014547

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. https://doi.org/10.1037/a0019339

Thompson, R. L., Vinson, D. P., Woll, B., & Vigliocco, G. (2012). The road to language learning is iconic: Evidence from British Sign Language. *Psychological Science*, *23*(12), 1443–1448. https://doi.org/10.1177/0956797612459763

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

Tomasuolo, E., Bonsignori, C., Rinaldi, P., & Volterra, V. (2020). The representation of action in Italian Sign Language (LIS). *Cognitive Linguistics*, *31*(1), 1–36. https://doi.org/10.1515/cog-2018-0131

Ungerer, F. (1999). Diagrammatic iconicity in word-formation. In M. Nänny & O. Fischer (Eds.), *Form miming meaning* (pp. 307-324). Amsterdam & Philadelphia: John Benjamins.

Vermeerbergen, M. (2006). Past and current trends in sign language research. *Language & Communication, 26(2)*, 168-192. https://doi.org/10.1016/j.langcom.2005.10.004

Vermeerbergen, M., Leeson, L., & Crasborn, O. A. (2007). *Simultaneity in signed languages: Form and function*. Amsterdam & Philadelphia: John Benjamins.

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 B: Biological Sciences*, *369*(1651), 20130292. https://doi.org/10.1098/rstb.2013.0292

Vigliocco, G., Vinson, D. P., Woolfe, T., Dye, M. W. G., & Woll, B. (2005). Language and imagery: Effects of language modality. *Proceedings of the Royal Society B: Biological Sciences*, *272*(1574), 1859–1863. https://doi.org/10.1098/rspb.2005.3169

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. https://doi.org/10.1016/j.jml.2015.03.002

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2001). Storage of features, conjunctions, and objects in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *27*(1), 92–114. https://doi.org/10.1037/0096-1523.27.1.92

Volterra, V. (1987). *LIS: La Lingua Italiana dei Segni*. Bologna: Il Mulino.

Volterra, V. (2011). La ricerca sulla lingua dei segni in Italia: Passato, presente e prospettive future. In A. Cardinaletti, C. Cecchetto, & C. Donati (Eds.), *Grammatica, lessico e dimensioni di variazione nella LIS*, 27–44.

Volterra, V., Roccaforte, M., Di Renzo, A., & Fontana, S. (2022). *Italian Sign Language from a cognitive and socio-semiotic perspective*. John Benjamins Publishing Company. https://doi.org/10.1075/gs.9

Wilbur, R. B. (1987). *American sign language: Linguistic and applied dimensions*. Boston: College-Hill Press

Wilbur, R. B., & Nolen, S. B. (1986). The duration of syllables in American sign language. *Language and Speech*, *29*(3), 263–280. https://doi.org/10.1177/002383098602900306

Wilcox, S. (2004). Cognitive iconicity: Conceptual spaces, meaning, and gesture in signed language. *Cognitive Linguistics*, *15*(2), 119–147. https://doi.org/10.1515/cogl.2004.005

Wilcox, S., & Occhino, C. (2016). Constructing signs: Place as a symbolic structure in signed languages. *Cognitive Linguistics*, *27*(3), 371–404. https://doi.org/10.1515/cog-2016-0003

Wilkinson, E. L. (2009). *Typology of signed languages: Differentiation through kinship terminology* [Doctoral dissertation, The University of New Mexico]. https://www.proquest.com/docview/304946043?pq-origsite=gscholar&fromopenview=true

Wilkinson, E. (2016). Finding frequency effects in the usage of NOT collocations in American Sign Language. *Sign Language & Linguistics*, 19(1), 82-123.

Wilson, J. (1996). The tobacco story: Narrative structure in an American Sign Language story. *Multicultural Aspects of Sociolinguistics in Deaf Community*, 152–182.

Wilson, M., & Emmorey, K. (1997). Working memory for sign language: A window into the architecture of the working memory system. *Journal of Deaf Studies and Deaf Education*, *2*(3), 121–130.

Wilson, M., & Emmorey, K. (2006). No difference in short-term memory span between sign and speech. *Psychological Science*, *17*(12), 1093–1094. https://doi.org/10.1111/j.1467-9280.2006.01835.x

Winston, E. A. (1991). Spatial referencing and cohesion in an American Sign Language Text. *Sign Language Studies*, *73*(1), 397–410. https://doi.org/10.1353/sls.1991.0003

Winston, E. A. (1992). Space and involvement in an American Sign Language lecture. In J. Plant-Moeller (Ed.), *Expanding horizons: Proceedings of the Twelfth National Convention of the Registry of Interpreters for the Deaf, August 6-11, 1991* (pp. 93–105). RID Publications.

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. (2006). ELAN: a professional framework for multimodality research. In *Proceedings of the 5th International Conference on Language Resources and Evaluation* (LREC 2006) (pp. 1556–1559). https://pure.mpg.de/pubman/faces/ViewItemOverviewPage.jsp?itemId=item_60436

Woodman, G. F., Vecera, S. P., & Luck, S. J. (2003). Perceptual organization influences visual working memory. *Psychonomic Bulletin & Review*, *10*(1), 80–87. https://doi.org/10.3758/BF03196470

Yamashita, H., & Chang, F. (2001). 'Long before short' preference in the production of a head-final language. *Cognition*, *81*(2), B45–B55. https://doi.org/10.1016/S0010-0277(01)00121-4

Zipf, G. K. (1949). *Human behavior and the principle of least effort: An introduction to human ecology*. Cambridge: Addison-Wesley Press.

Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Language comprehenders mentally represent the shapes of objects. *Psychological Science*, *13*(2), 168–171. https://doi.org/10.1111/1467-9280.00430

Zwaan, R. A., & Yaxley, R. H. (2003). Spatial iconicity affects semantic relatedness judgments. *Psychonomic Bulletin & Review*, *10*(4), 954–958. https://doi.org/10.3758/BF03196557

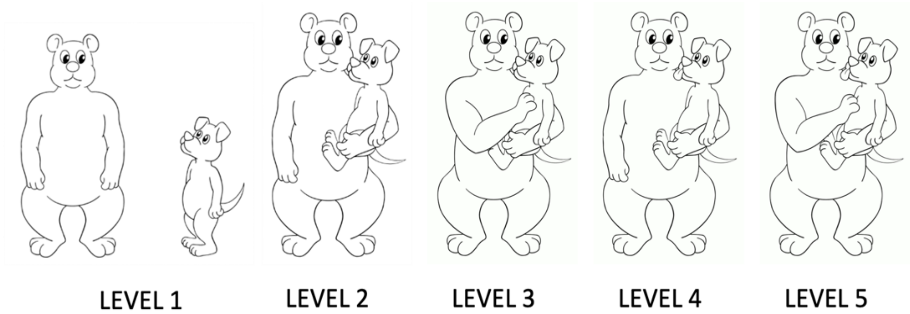
Zwitserlood, I. (2003). *Classifying hand configurations in Nederlandse Gebarentaal* [Doctoral dissertation, Radboud University]. https://www.lotpublications.nl/classifying-hand-configurations-in-nederlandse-gebarentaal-classifying-hand-configurations-in-nederlandse-gebarentaal

Appendix A

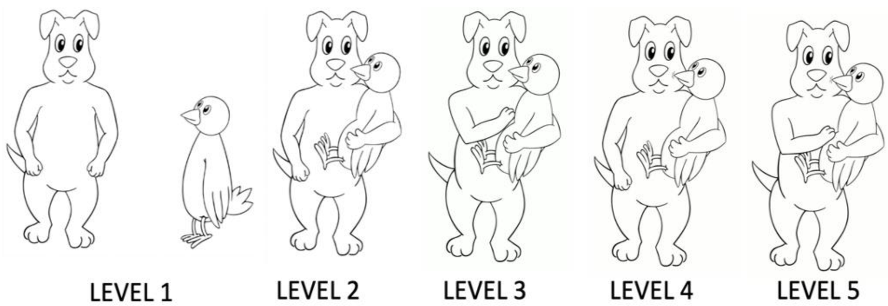
Stimuli sets used in the experiments

Note that figures in levels 3, 4 and 5 are GIFs and dynamic actions are animated). All images in their original format are freely available online (https://osf.io/g57p2/).

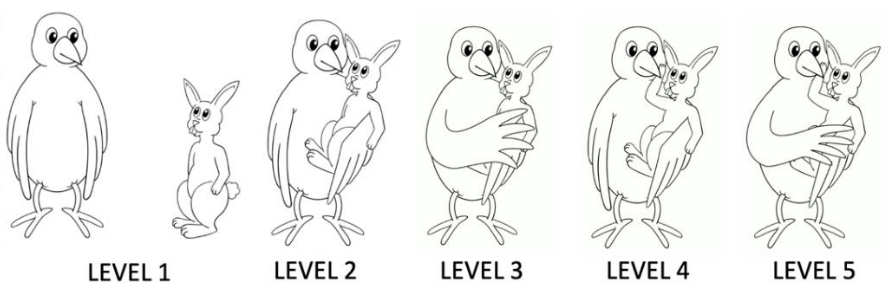
Set: ***Bear & Dog*** (Dynamic action 2 – *licking*):



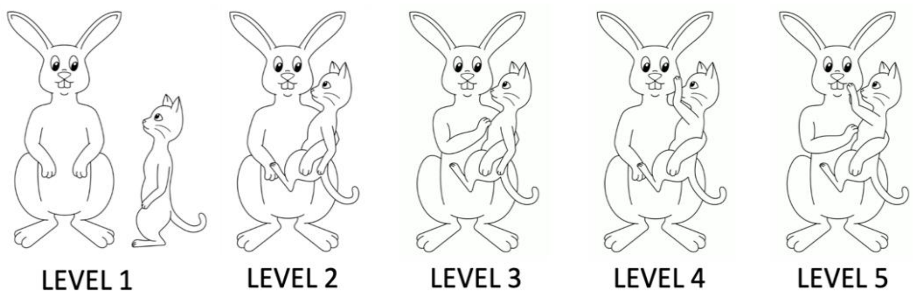
Set: ***Dog & Bird*** **(**Dynamic action 2 – *pecking*):



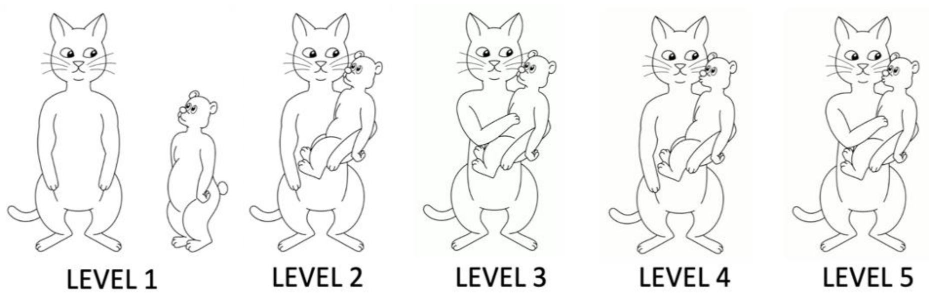
Set: ***Bird & Bunny*** (Dynamic action 2 – *tapping*):



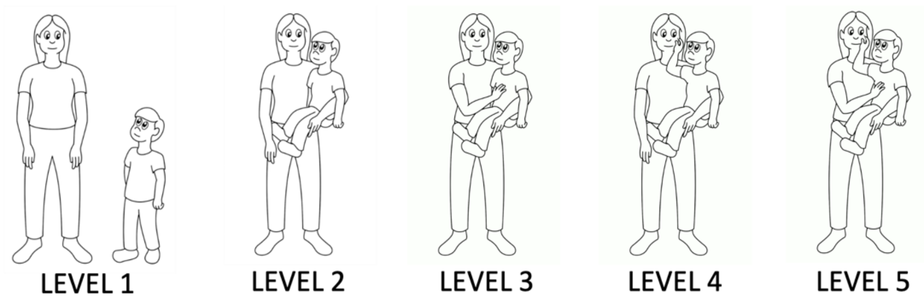
Set: ***Bunny & Cat*** (Dynamic action 2 –*petting*):



Set: ***Cat & Bear*** **(**Dynamic action 2 – *kissing*):



Set: ***Woman & Child.*** Dynamic action 2 – *pinching.*



Appendix B

Research Data Management

Use of existing data:

No existing data were used

New data:

Two sets of new data were created:

Twenty-three deaf adults (citizens of Italy, LIS signers) were video-recorded in Rome, Italy at the Institute of Cognitive Science and Technologies, National Research Council Rome. Data were used in Chapters 2, 3 and 4 of the thesis.

Twenty-three hearing adults (citizens of Italy, Italian speakers) were video-recorded in Rome, Italy at the Institute of Cognitive Science and Technologies, National Research Council Rome. Data were used in Chapter 4 of the thesis.

Privacy issues (in terms of ethical permissions and GDPR requirements):

All participants signed informed consent forms (it. *consenso informato*) regarding the use of the video recorded data and the personal data they provided in the questionnaire.

Ethical permission/consent form provided permission to use the data of the study, anonymously, to process and archive them so that they can be used for future research at the CNR Rome in compliance with the Italian law 196/2003 on the confidentiality of personal data and the European regulation on the processing of personal data of individuals (GDPR or General Data Protection Regulation). Consent forms provided an option for a participant to consent to use of their video-recording or screenshots of the video-recording for academic presentations, academic journals and education purposes.

Following personal data were collected:

Video recordings of the participants

Following personal data of deaf Italian participants were collected by means of a questionnaire:
Date of birth, sex, handedness, other deaf persons in the family, language/s used in family, age of acquisition of LIS, place of acquisition of LIS, level of deafness, use of protheses, use of cochlear implant.

Following personal data of hearing Italian participants were collected by means of a questionnaire:
Sex, handedness, language/s used in the family, age of acquisition of Italian, knowledge of other languages.

Personal data regarding the name, surname, e-mail address, and phone number were collected for administrative purposes and were not used as research data and excluded from the research data material.

If data can be shared for re-use with others, where and under which access conditions the data can be accessed.
Raw data (i.e., video recordings) cannot be shared for re-use.
Processed anonymized data can be accessed through OSF repository:
Chapter 2 (LIS data regarding simultaneous information encoding): https://osf.io/mwg4v/
Chapter 3 (LIS data regarding linguistic strategy use): https://osf.io/zwex8/
Chapter 4 (LIS data regarding simultaneous information encoding and silent gesture (hearing participants) data regarding simultaneous information encoding): https://osf.io/uw2jd/

Nederlandse samenvatting

Een fundamentele aanname over taal – of het nu gesproken of visuele taal betreft – is dat er sprake is van een linearisatie probleem: in de wereld gebeuren veel dingen tegelijkertijd, maar als we over een gebeurtenis willen vertellen kunnen we niet al die elementen tegelijkertijd uitdrukken. In gesproken taal wordt dat meestal als volgt opgelost: verschillende elementen van een gebeurtenis worden opgedeeld in verschillende taalonderdeeltjes, zodat ze vervolgens één voor één achter elkaar kunnen worden geplaatst (een lineaire rangschikking). Maar in gebarentalen – waar gebaren zichtbaar in plaats van hoorbaar zijn -, wordt betekenis niet enkel lineair uitgedrukt, maar ook simultaan. Meerdere elementen van een gebeurtenis kunnen tegelijkertijd worden uitgedrukt door middel van iconiciteit en het gebruik van meerdere lichaamsdelen (handen, torso, hoofd, gezicht en ogen). Dit biedt communicatieve voordelen: door gebruik te maken van gelijktijdigheid en iconiciteit (eigenschappen die inherent zijn aan de modaliteit van gebarentaal), kunnen complexe gebeurtenissen op een efficiënte manier worden naverteld. Maar alles wat we tot nu toe weten over communicatieve efficiëntie komt van onderzoek naar gesproken talen. Daardoor weten we nauwelijks iets over of en hoe de modaliteit van een taal een rol speelt als het gaat om communicatieve efficiëntie – een kernfunctie van taal. Het algemene doel van dit proefschrift was daarom om vast te stellen of gelijktijdigheid en iconiciteit gebruikt worden in het LIS (Italiaanse gebarentaal) om efficiënt te communiceren, en of deze eigenschappen zijn ontstaan om efficiëntie te optimaliseren.

Deze vragen zijn onderzocht door middel van drie experimentele studies, die verschillende maar complementaire perspectieven bieden op de rol van gelijktijdigheid en iconiciteit als het gaat om de organisatie van informatie (hoofdstuk 2), talige strategieën om informatie uit te drukken (hoofdstuk 3) en taalevolutie (hoofdstuk 4). Dit proefschrift gaat verder dan de bestaande literatuur, niet enkel op theoretisch maar ook op empirisch vlak. Een nieuw type experiment is ontwikkeld en geïmplementeerd, waarmee communicatieve efficiëntie gemeten kan worden in interacties over gebeurtenissen met verschillende niveaus van complexiteit.

Door onderzoek over gebarentaal, communicatieve efficiëntie en taalevolutie samen te brengen, laat dit proefschrift zien dat een focus op gebarentaal nieuwe inzichten op kan leveren over de algemene en modaliteit-specifieke strategieën die taalgebruikers benutten voor communicatieve efficiëntie. Bovendien vestigt dit proefschrift de aandacht op de functionaliteit van zowel lineaire als simultane linguïstische structuren, en draagt het bij aan de groeiende kennis over iconiciteit als een van de kerneigenschappen van taal.

Acknowledgments

This PhD journey was not a usual one. It spanned three countries, a global pandemic, and an outbreak of war. Thus, although pursuing a PhD was a dream come true for me, this experience was also full of turmoil. For this reason, I am not sure I will be able to find enough words to express my gratitude to the people that helped me navigate this journey in such extraordinary times. However, I will try to do my best.

First and foremost, thank you to the two brilliant women that I was lucky to have as my supervisors, Prof. dr. Asli Özyürek and dr. Olga Capirci. This was a real dream team and I could not wish for better mentors.

Olga, my journey could start because you trusted me to be a suitable candidate for the DCOMM project. You knew that I had very little knowledge about the sign language field but you could also see how passionate I was to delve deep into it. Thank you for letting my enthusiasm turn into this thesis. Our lively conversations about *what language is* still remain some of the fondest memories I have from Rome. You encouraged me to look for the truth beyond the mainstream. I am forever grateful for that. Thank you for trusting me to pursue my ideas and providing support whenever I needed it. In doing so, you helped me find my confidence as a researcher. Thank you for contributing to shaping me into the researcher I am now.

Asli, we first met when I knew nothing about research, and now, here I am, writing acknowledgments of my PhD thesis. Thank you for being my mentor from the very beginning. Thank you for accepting me as an external-PhD and for finding a way for me to return in Nijmegen so I could enjoy the stimulating environment of the MLC lab and your guidance in person. I would not be at this final stage now if not for you. Thank you for always having time for me. Thank you for all the inspiring discussions in your office, in hallways of MPI, during lunch, or on our way to talks and lectures. They made me a better researcher and they opened my mind to new ideas and aspirations. Thank you for sharing your wisdom so I could become a stronger woman in academia and outside of it. And finally, thank you for believing in me and giving me an opportunity to continue this adventure. I am excited for what is yet to come.

Thank you to my manuscript committee—Prof. dr. Onno Crasborn, Prof. dr. Edward Gibson, Prof. dr. Pamela Perniss, Dr. Erin Wilkinson, and Dr. Sabina Fontana—for the time devoted to reading this thesis. I am lucky to have my thesis evaluated by people whose work I admire.

I would like to express an immense thank you to the deaf community in Rome and especially to Gruppo Silis. Thank you for not only teaching me LIS but also introducing me to and accepting me into the community. Thank you for giving me an opportunity to learn and to grow from within and not just as an observer. A very special thank you goes to the most wonderful teachers of LIS—Simona, Laura, and Betty. Non ho mai avuto le insegnanti più brave di voi. Esattamente grazie a voi mi sono innamorata della LIS. Grazie mille di tutto, grazie dell'entusiasmo, della sincerità e della pazienza!

The research described in this thesis would have never been possible without the participants, deaf and hearing, that took part in the studies. Thank you so much for your time. Thank you to the student assistants who helped with coding the data—Alessandra, Chiara, and Giulia. I sincerely hope that all of you will be pleased to see the fruits of your participation.

Thank you to the labs that welcomed me during my PhD journey—LaCAM lab in Rome where it all started and MLC lab, now Multimodal Language Department, in Nijmegen where this journey is approaching its end.

Thank you to the past and present LaCAM members—Olga, Alessio, Barbara, Chiara, Cristiana, Cristina, Elena, Francesca, Laura, Luca, Morgana, Pasquale, Stefano, Tommaso, and Virginia. Thank you for all your support and advice. Thank you for all the help with recruiting participants and running experiments. Ci sono stati quattro anni a Roma bellissimi e senza di voi i giorni lavorativi non sarebbero stati cosi stimolanti e divertenti! Grazie per avermi accettato nel vostro gruppo, per avermi accettato con tute le mie stranezze lettoni che, immagino, potrebbero a volte sembrare troppo fredde per i cuori caldi italiani. Vi apprezzo molto. Grazie per aver reso Roma un posto molto speciale per me. E grazie per aver sopportato il mio italiano! Thank you to Morgana and Chiara, my Roman partners in crime, for all the fun times and for making me feel welcome in the Eternal City.

Thank you to the past and present MLC members that I was lucky to meet, would it be in person or via Zoom—Asli, Andrea, Beyza, Christina, Dilay, Erce, Ezgi, Francie, Gerardo, Hükümran, Hüma, James, Jessica, Jing, Judith, Kazuki, Linda, Louise, Madeleine, Mark, Marlou, Martina, Pelin, Renske, Sevilay, Sho, Susanne, Tom, Uğur, Vicky, Wim, and Zeynep. Many of you witnessed the most intense period of not only my PhD journey but my life so far in general. Thank you for your support and understanding this past year. Thank you for stimulating discussions and feedback. I am happy to be a part of the group that consists of such kind and bright people. I am looking forward to learning and growing together.

A special thank you to my paranymphs, Ezgi & Marlou. Thank you for keeping me sane during these last months and reminding me that the toughest part of this journey is already over. I am lucky to have you by my side as I am preparing for the defence.

Thank you to the DCOMM members and especially to my fellow ESRs—Baris, Emanuela, Gabriella, Mehmet, Merlijn, Michela, Raed, Roberta, Samuel, Sara, and Yesol. We embarked on this journey not really realizing what pursuing a PhD degree really means and I am glad that together we managed to find a balance between intense training and having fun. Roberta, I am glad DCOOM brought us together. Thank you for all the shared experiences.

I am very grateful to my dearest Latvians—Anna Petra, Austra, Liene, Lana, Laura, and Renāte. You cheered me up when I was down and celebrated my victories. Thank you for being proud of me although you never fully knew what it is I actually do. Pursuing a PhD is a serious thing, but you added a bubbly touch to the entire experience

Ezgi, I don't even know where to start. I came to Nijmegen to finish my PhD, bracing for the months of emotional solitude. But you came into my life—a blessing I have never hoped for. Without you, this journey would have been so much harder and so much lonelier. I have way too many things to thank you for—encouragement, support, advice, for looking out for me and for being the greatest ally in fighting my imposter syndrome. But most of all, thank you for your kindness and sincerity. Thank you for being there for me. I am so happy to have a friend like you.

Tom, this experience was hard for us on a whole different level. Thank you for letting me go so I can follow my dream, never once doubting my love for you. Paldies, ka esi bijis līdzās arī brīžos, kad bijām šķirti. Paldies, ka iedvesmo un iedrošini cīnīties par sapņiem. Paldies, ka tici maniem spēkiem un vienmēr atbalsti.

Finally, I owe the biggest thanks to my extraordinary parents. Thank you for encouraging me to make my own choices in life. Thank you for doing everything possible to support me as I was exploring my passions. I hope this thesis shows that you were not mistaken in letting me follow my heart. Paldies, ka iemācījāt tiekties pēc zināšanām. Tā ir dāvana, kuras vērtību sāku patiesi apjaust tikai tagad. Paldies, mammu. Спасибо, папа.

Curriculum Vitae

Anita Slonimska was born in 1990 in Riga, Latvia. In 2013, she obtained her Bachelor's degree in Italian Culture from Latvian Academy of Culture, Latvia. In 2015, Anita obtained her Master's degree in Linguistics, and in 2016, her second Master's degree (research) in Language and Communication with a specialization in Psycholinguistics (*cum laude*) from Radboud University, Nijmegen, the Netherlands.

In 2016, Anita began her PhD training within the European Union's Horizon 2020 research and innovation program DCOMM under the Marie Sklodowska-Curie Actions Grant in the Language and Communications Across Modalities (LaCAM) group at the Institute of Cognitive Sciences and Technologies, National Research Council of Rome, and in the Multimodal Language and Cognition (MLC) group at the Centre for Language Studies, Radboud University.

Currently, Anita is a postdoctoral researcher in the Multimodal Language Department at the Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands.

Publications

Slonimska, A., Ozyurek, A., & Capirci, O. (2022). Simultaneity as an emergent property of sign languages. In A. Ravignani, R. Asano, D. Valente, F. Ferretti, S. Hartmann, M. Hayashi, et al. (Eds.), *The evolution of language: Proceedings of the Joint Conference on Language Evolution (JCoLE)* (pp. 678-680). Nijmegen: Joint Conference on Language Evolution (JCoLE).

Slonimska, A., Özyürek, A., & Capirci, O. (2022). Simultaneity as an emergent property of efficient communication in language: A comparison of silent gesture and sign language. *Cognitive Science, 46*(5), e13133.

Slonimska, A., Özyürek, A., & Capirci, O. (2021). Using depiction for efficient communication in LIS (Italian Sign Language). *Language and Cognition, 13*(3), 367-396.

Slonimska, A., Özyürek, A., & Capirci, O. (2020). The role of iconicity and simultaneity for efficient communication: The case of Italian Sign Language (LIS). *Cognition, 200,* 104246.

Slonimska, A., & Roberts, S. G. (2017). A case for systematic sound symbolism in pragmatics: Universals in wh-words*. Journal of Pragmatics, 116*, 1-20.

Slonimska, A., & Roberts, S. G. (2017). A case for systematic sound symbolism in pragmatics:The role of the first phoneme in question prediction in context. In G. Gunzelmann, A. Howes, T. Tenbrink, & E. Davelaar (Eds.*), Proceedings of the 39th Annual Conference of the Cognitive Science Society (CogSci 2017)* (pp. 1090-1095). Austin, TX: Cognitive Science Society.

Slonimska, A., Özyürek, A., & Campisi, E. (2015). Ostensive signals: markers of communicative relevance of gesture during demonstration to adults and children. *Proceedings of the 4th GESPIN-Gesture & Speech in Interaction Conference* (pp. 217-222). Nantes: Universite of Nantes.