COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 46 (2022) e13133

ISSN: 1551-6709 online
DOI: 10.1111/cogs.13133

Simultaneity as an Emergent Property of Efficient Communication in Language: A Comparison of Silent Gesture and Sign Language

Anita Slonimska,[a,b] Asli Özyürek,[a,b,c] Olga Capirci[d]

[a] *Centre for Language Studies, Radboud University, Nijmegen, The Netherlands*
[b] *Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands*
[c] *Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, the Netherlands*
[d] *Institute of Cognitive Sciences and Technologies (ISTC), National Research Council (CNR) of Italy*

Received 23 August 2020; received in revised form 25 February 2022; accepted 16 March 2022

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. 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

---

Correspondence should be sent to Anita Slonimska, Radboud University, Centre for Language Studies, Erasmusplein 1, 6525 HT Nijmegen, The Netherlands. E-mail: Anita.Slonimska@ru.nl

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.

Keywords: Efficient communication; Language emergence; Simultaneity; Iconicity; Sign language; Silent gesture; Visual modality

---

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, McNeill, & Singleton, 1996; Kirby, Cornish, & Smith, 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 (e.g., Kirby et al., 2008; Motamedi, Schouwstra, Smith, Culbertson, & Kirby, 2019; Senghas, Kita, & Özyürek, 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, 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, Özyürek, & Capirci, 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, 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, Lucero, & Goldin-Meadow, 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, So, Özyürek, & Mylander, 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.

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, Furman, & Goldin-Meadow, 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, Kirby, & Brighton, 2003, p. 372) into segmented and compositional structure through conventionalization of the segmented meaning elements and their systematic recombination to create new meanings (Beckner, Pierrehumbert, & Hay, 2017; Kirby, 2000; Kirby et al., 2008, 2014, 2015; Motamedi et al., 2019; Nölle, Staib, Fusaroli, & Tylén, 2018; Raviv, Meyer, & Lev-Ari, 2019; Theisen, Oberlander, & Kirby, 2010). These studies show that this process emerges as an adaptation to pressures of language use and transmission to new learners, which push languages toward 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 Nicaraguan Sign Language (NSL) 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 signers, 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.

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 and 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, that is, 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, Herman, & Woll, 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, we still know practically nothing about the emergent trajectory of the simultaneous use of multiple articulators to encode distinct semantic elements in an event. 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, Senghas, and Snedeker (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.

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, & Sevcikova-Sehyr, 2015; Cuxac, 1999, 2000; Napoli & Sutton-Spence, 2010; Perniss, 2007; Quinto-Pozos, 2007; Risler, 2007). Iconicity does not only refer to the resemblance between a single linguistic form and its meaning, that is, *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, that is, *diagrammatic iconicity* (Perniss, 2007; Risler, 2007; Slonimska, Özyürek, & Capirci, 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; Vermeerbergen, Leeson, & Crasborn, 2007). However, signers use not only both manual articulators in a simultaneous manner but also nonmanual 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, for example, 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 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 depict the respective body parts of the referent, and the facial expression of the signer depicts the emotion expressed by 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 (Fig. 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, that is, placing one hand above the hand holding the bear and depicting a petting motion (Fig. 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).

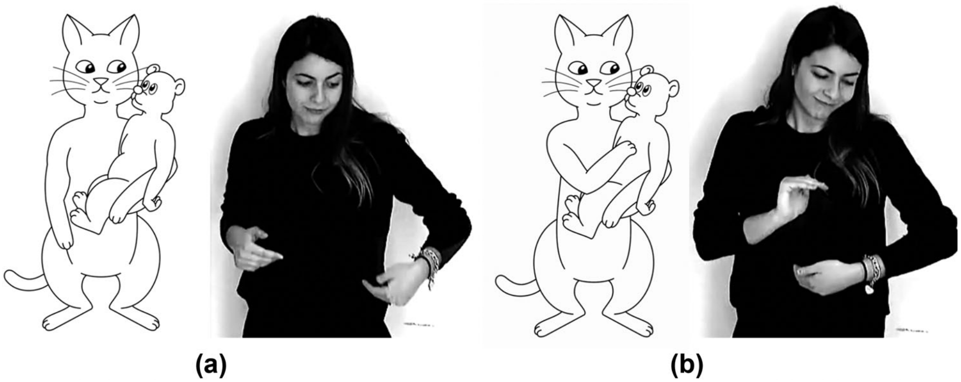

Fig. 1. A signer encoding (a) a cat holding and looking at the bear in its hand, and (b) a cat holding, looking at and petting the bear in its hand.

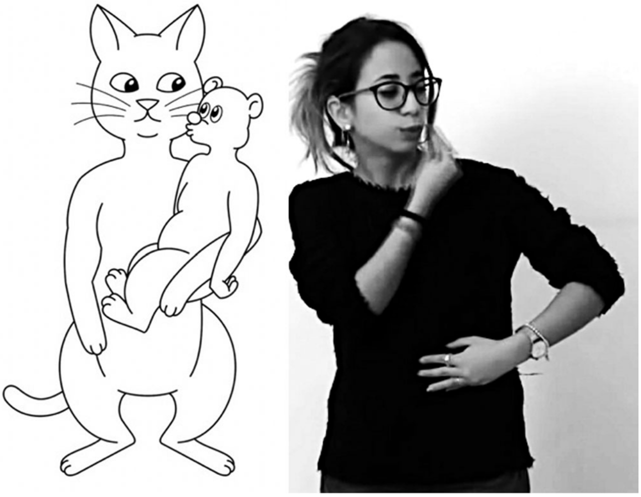

Fig. 2. A signer encoding a cat holding and looking at the bear in its hand and the bear kissing the cheek of the cat.

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, that is, 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 (Fig. 2), a signer can use lexical sign for *kissing* by positioning this sign in a diagrammatic relation to the recipient of the kiss, that is, 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.

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, that is, 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. 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. 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. 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. 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, that is, 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.

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.

3. Method

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.

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.

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, and dynamic action of referent 2 (see Fig. 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);

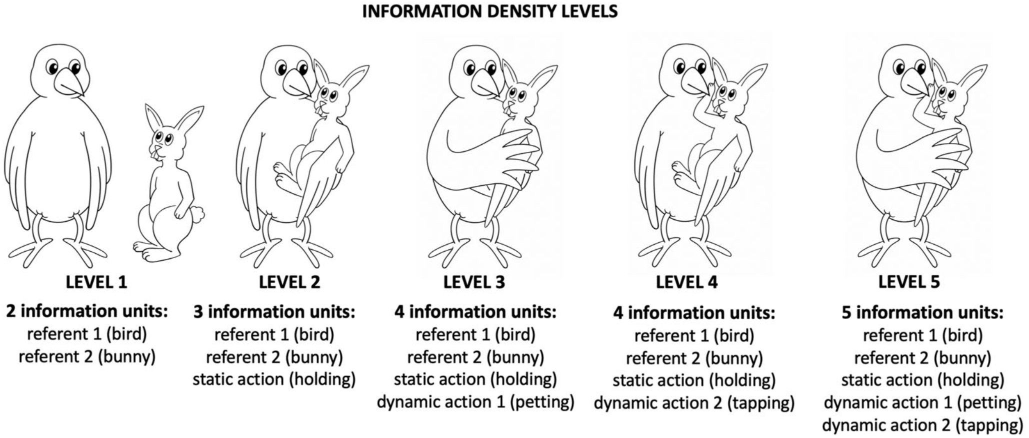

Fig. 3. PNG items for level 1 (a) and level 2 (b), GIF items for level 3 (c), level 4 (d), and level 5 (e). In the GIFs, the dynamic action of referent 1 (petting) and dynamic action of referent 2 (tapping) are animated. All original items can be found online (https://osf.io/g57p2).

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.

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 nonverbal 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 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.

3.4. Coding

The video-recorded data was coded in multimodal data annotation software ELAN, developed by Max Planck Institute for Psycholinguistics (Wittenburg, Brugman, Russel,

	MS1a	MS1b	MS2	MS3	MS4a	MS4b	MS5
	(Hand gesture)	*(Hand gesture)*	*(Hand gesture)*	*(Movement)*	*(Hand gesture)*	*(Hand gesture)*	*(Non-manual gest.)*
RH:	(*cat*)	(*cat*)	(*holding*)	-	(*bear*)	(*bear*)	-
LH:	(*cat*)	(*cat*)	-	-	(*bear*)	(*bear*)	-
Non-manuals:			(torso - *dog)*				(torso - *bear;* face - *kiss*)
Movement:				(*bear)*	(*bear*)	(*bear*)	(*bear)*
Kinem. Simult:	no	*NA*	yes	no	no	*NA*	yes
N. of info:	1	*NA*	2	1	1	*NA*	2

Fig. 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.

Klassmann, & Sloetjes, 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 nonmanual 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.*

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, nonmanual gesture*, or *whole-body movement*. Movement segments that included manual gesture strokes (with or without use of nonmanual articulators and/or body movement) were classified as *manual gestures*, while movement segments produced only by nonmanual articulators were classified as *nonmanual 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, MS5). In addition, silent gesturers used whole body movements, stepping to the right/left space and remaining there to encode referents (see Fig. 4, MS3, MS4a, MS4b, MS5). If they were used in absence of manual gestures or nonmanual gestures, a movement segment was coded as *whole-body movement*, as in Fig. 4, MS3 (as opposed to *manual* or *nonmanual gesture*). If the whole-body movements were used simultaneously with manual or nonmanual gestures, they were classified accordingly—*manual gesture* in the former case (Fig. 4, MS4a, MS4b) and *nonmanual gesture* in the latter case (Fig. 4, MS5). Note that in Slonimska et al. (2020) signers never used nonmanual 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 nonmanuals simultaneously), for silent gesturers, movement segments could be also *nonmanual 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 MS1a and MS1b; a gesture for *ears* followed by a gesture for *cheeks* to refer to *a bear*, see Fig. 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. each of the encodings for *cat* and *bear* consisted of 2 gestures to form a compound name for the referent (MS1a and MS1b for *cat*, and MS4a and 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).

3.4.2. Kinematic simultaneity

For each movement segment we coded whether articulators (manual and nonmanual: 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) the gesturer uses both hands to encode the cat—referent 1 (Fig. 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, MS5, Movement) together with use of the torso encoded referent 2 – the *bear* (see Fig. 4, MS5, Nonmanuals), while the facial expression, or more specifically the gesturer's mouth, encoded the action of kissing (see Fig. 4, MS5, Nonmanuals). 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.

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, *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 nonmanual 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, 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 (Fig. 5).

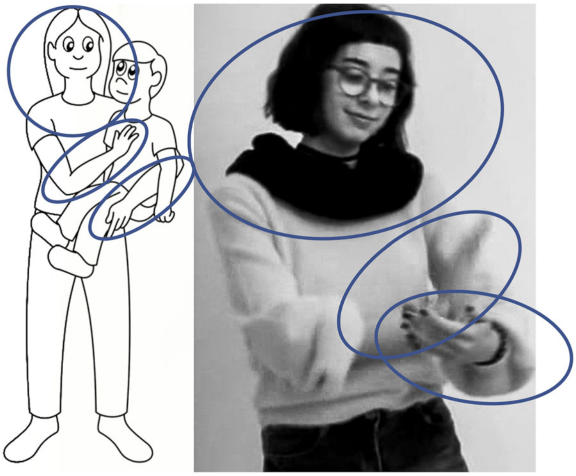

Fig. 5. An example of a movement segment (MS) with three information units. Blue circles represent semantic information units of the same referent (the woman).

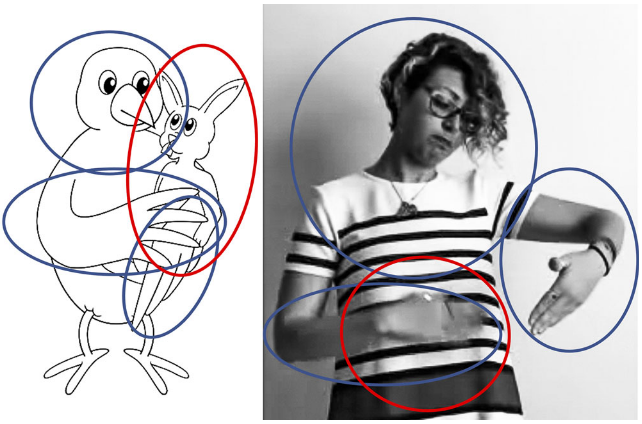

Fig. 6. An example of a movement segment (MS) with four information units. Colored circles represent semantic information units of different referents (blue circles for the bird, red circles for the bunny).

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 (Fig. 6).

3.5. Reliability

Data were 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 .97. Reliability for movement segment type (*manual gesture, nonmanual gesture, whole-body movement*) between coders was very strong (*Cohen's* κ =.89). Reliability for simultaneous use of multiple articulators (manual and nonmanual) in each movement segment (*Cohen's* $\kappa = .91$) as well as reliability regarding number of simultaneously encoded information units in each movement segment were also very strong (*Cohen's* κ =.86).

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.

4.1. 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*, that is, a movement where at least one hand was used to perform a stroke, 270 were classified as *nonmanual gestures*, that is, 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, Levy, Scheepers, & Tily, 2013). For all independent variables analyzed, that is, *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/).

4.1.1. 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 Fig. 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.

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* ($\chi^2(1) = 1.54$, $p = .214$)*, age* ($\chi^2(1) = 0.79$, $p = .374$) and *handedness* ($\chi^2(1) = 0.004$, $p = .950$). 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 ($\chi^2(4) = 70.41$,

$p < .001$, see Table 1 for a summary of the model). The contrasts between levels shown in Table 1 indicate a significant gradual increase in length of productions, except levels 3 and 4, which were comparable.

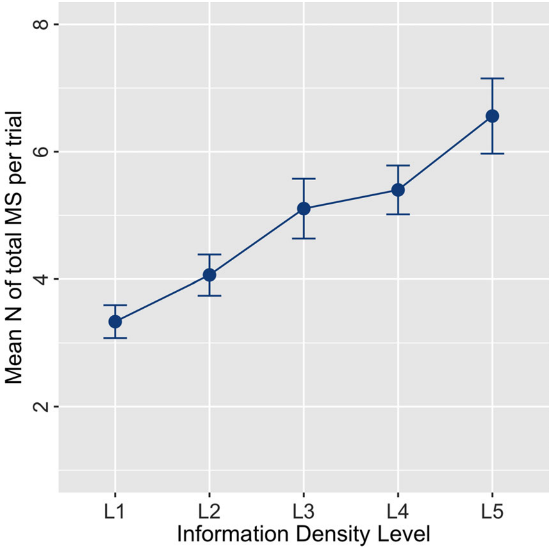

Fig. 7. Means of total number of movement segments (MS) used per response to a trial by silent gesturers. Error bars indicate a 95% CI for observations grouped within participants.

Table 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.[1]

Random effects	**Variance**	***SD***				
Trial	0.00	0.00				
Participant	0.008	0.088				
Number of obs: 647	Trial = 30	Participant = 23				
Fixed effects		95% CI		*SE*	*z* Value	*p* Value
	β	Lower b	Upper b			
(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

[1] 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.

4.1.2. *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 (Fig. 8). In level 1, use of simultaneity was close to nonexistent ($M = 0.007$, $SD = 0.02$) 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 ($M = 0.46$, $SD = 0.12$).

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* ($\chi^2(1) = 0.86$, $p = .35$), *age* ($\chi^2(1) = 0.30$, $p = .59$), and *handedness* ($\chi^2(1) = 1.02$, $p = .31$) 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 ($\chi^2(4) = 122.37$, $p < .001$, see Table 2 for a summary of the model). The contrasts between levels shown in Table 2 indicate a significant gradual increase in kinematic simultaneity, except levels 3 and 4, which were comparable.

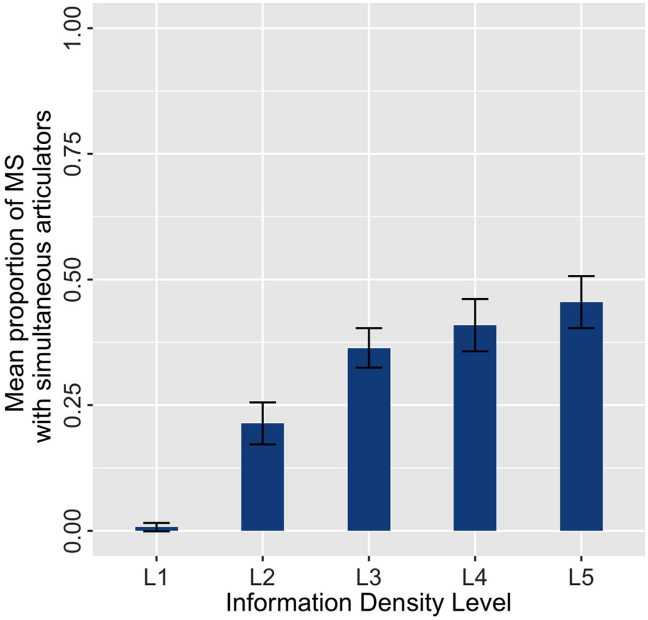

Fig. 8. Mean proportions of kinematically simultaneous movement segments (MS) out of total number of movement segments per trial. Error bars indicate 95% CI of observations grouped within participants.

Table 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

Random effects	Variance	*SD*				
Trial	0.00	0.00				
Participant	0.15	0.39				
Number of obs: 647	Trial: 30	Participant = 23				
Fixed effects		95% CI		*SE*	*z* Value	*p* Value
	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

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 seven movement segments encoded four information units (Fig. 9). However, six of the seven 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.

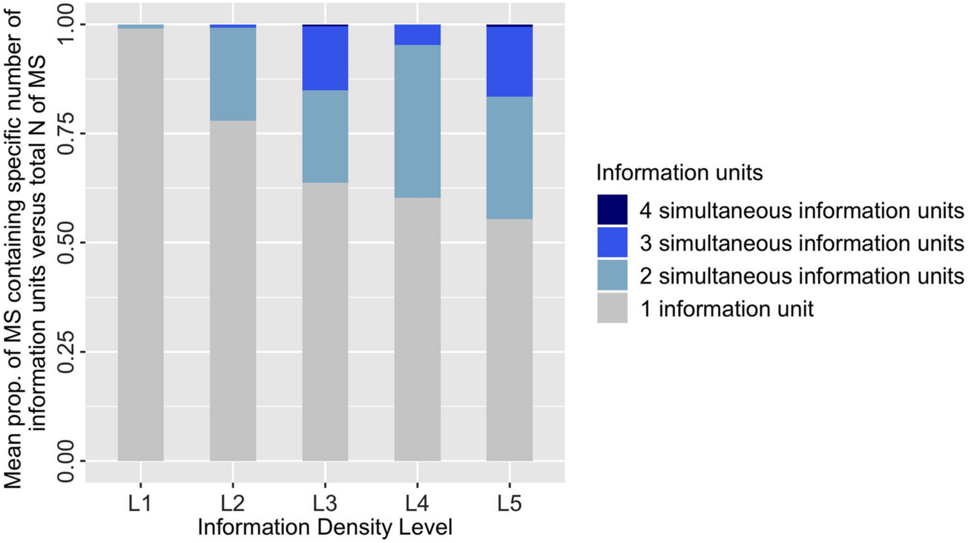

Fig. 9. *Mean* proportions of movement segments (MS) with 1, 2, 3, and 4 simultaneous information units out of the total number of movement segments in the silent gesture group.

Table 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

Random effects	Variance	SD				
Trial	0.00	0.00				
Participant	0.002	0.05				
Number of obs: 3145	Trial: 30	Participant = 23				
Fixed effects		95% CI		*SE*	*z* Value	*p* Value
	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

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* ($\chi^2(1) = 0.96$, $p = .32$), *age* ($\chi^2(1) =$ 8e-04, $p = .98$), and *handedness* ($\chi^2(1) = 1.02$, $p = .31$) 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 ($\chi^2(4) = 67.19$, $p < .001$, see Table 3 for a summary of the model). The contrasts between levels shown in Table 3 indicate a significant gradual increase in information density of simultaneity, except levels 3 and 4, which were comparable.

4.2. Silent gesture versus 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.

4.2.1. Length of encoding

We ran a series of Poisson mixed models in order to assess the effect of *Group* (silent gesture/ sign language) on the length of the productions (Table 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 ($\chi^2(4) = 95.76$, $p < .001$). 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*

Table 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

Models	Df	AIC	BIC	logLik	Deviance	χ^2	Df	*p* Value
Baseline	3	5234.8	5250.4	−2614.4	5228.8			
Information 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 × Group	12	5149.8	5212.0	2562.9	5125.8	2.16	4	.706

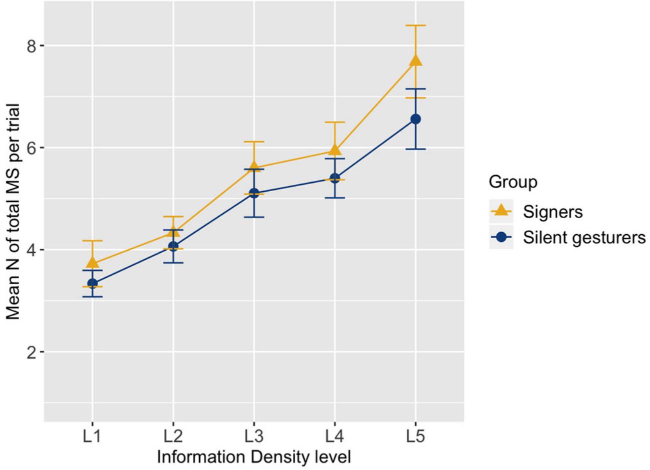

Fig. 10. Means of total number of movement segments (MS) used per each trial for silent gesture and sign language groups. Error bars indicate 95% CI of observations grouped within participants.

($\chi^2(1) = 5.12$, $p = .024$) but not the interaction ($\chi^2(4) = 2.16$, $p = .71$). The *Group* effect revealed that overall, the productions by silent gesturers were significantly shorter ($\beta = -0.10$, CI [−0.19; −0.02], $SE = 0.04$, $z = -2.33$, $p = .02$) than the productions by signers (Fig. 10).

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 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 ($\chi_2(4) = 140.57$, $p < .001$). 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* ($\chi^2(1) = 37.83$, $p < .001$) while the addition of the interaction did not reach the level of significance (see Table 5). The *Group* effect revealed that overall the productions by silent gesturers were

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

Models	Df	AIC	BIC	logLik	Deviance	χ^2	Df	*p* value
Baseline	3	3078.5	3094.1	−1536.3	3072.5			
Information 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 × Group	12	2908.8	2971.1	−1442.4	2884.8	9.33	4	.053

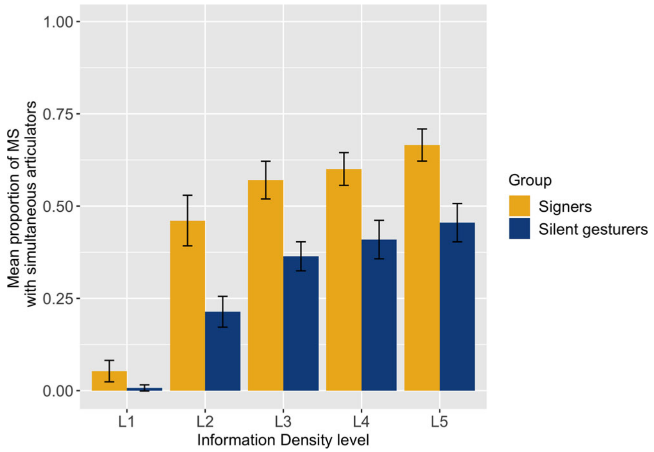

Fig. 11. Mean proportions of kinematically simultaneous movement segments (MS) out of total number of movement segments per trial for silent gesture and sign language groups. Error bars indicate 95% CI of observations grouped within participants.

significantly less simultaneous ($\beta = -0.98$, CI[-1.23; -0.73], $SE = 0.13$, $z = -7.65$, $p < .001$) than the productions by signers (Fig. 11).

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 6).

Table 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

Models	Df	AIC	BIC	logLik	Deviance	χ^2	Df	*p* Value
Baseline	3	17928	17948	−8961.0	17922			
Information density	7	17833	17881	−8909.7	17819	102.69	4	**<.001**
Group	8	17788	17842	−8885.7	17772	47.91	1	**<.001**
Info.density × Group	12	17787	17869	−8881.4	17763	8.59	4	.072

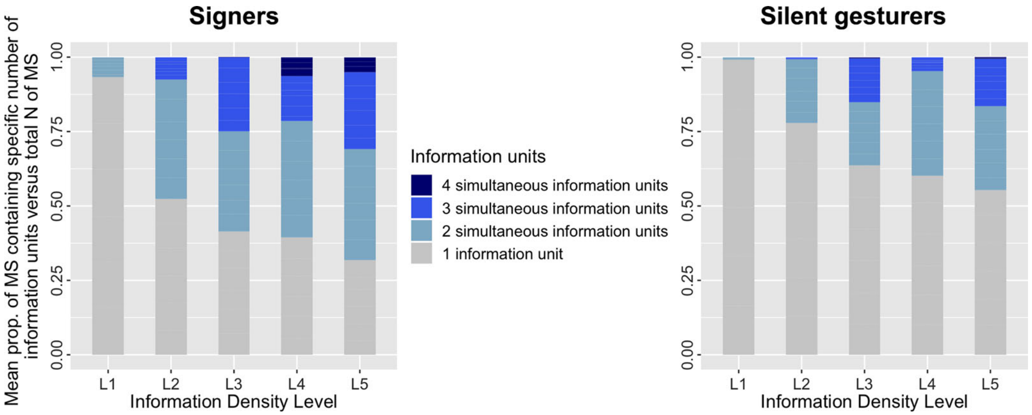

Fig. 12. Mean proportions of movement segments (MS) with 1, 2, 3, and 4 simultaneous information units out of the total number of movement segments in the sign language and silent gesture groups.

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 ($\chi^2(4) = 102.69$, $p < .001$). 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 ($\beta = -0.21$, CI[-0.26; -0.17], $SE = 0.02$, $z = -9.16$, $p < .001$) than productions by signers (Fig. 12).

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 Fig. 13), limiting their expressive possibilities. In contrast, signers always performed such actions with their hands, that is, mapping mouth-related movements onto their hands, with or without also mapping the action on the mouth (see Fig. 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

Fig. 13. Prototypical encoding by silent gesturers of dynamic action 2, which is performed by the mouth/head of the referent. Gesturers use their mouth/head to encode this action.

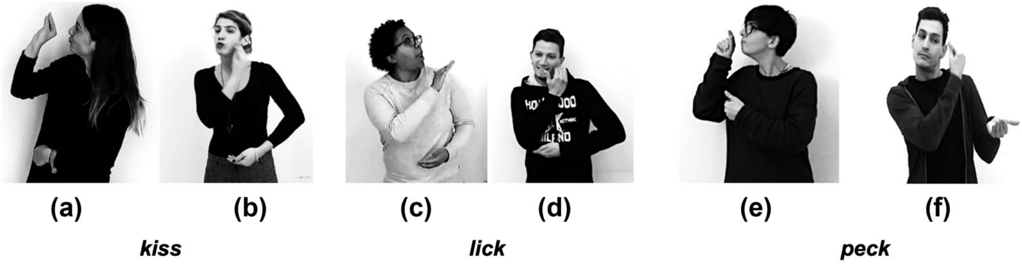

Fig. 14. Prototypical encoding strategies by signers of dynamic action 2, which is performed by the mouth/head of the referent. Signers use their hand, and additionally their mouth/head in some instances, to encode this action.

1 in a single simultaneous construction (Fig. 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.

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 eight used their hand to encode the action of pecking; the majority used their head/mouth (direct mapping). Only three silent gesturers used their hand to encode the action of licking (one of the eight who also encoded pecking) while none of the silent gesturers encoded the action of kissing with their hand (see Table 7).

However, in some instances direct mapping of the action could be taken advantage of, that is, 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 (Fig. 15), which allowed them to split this action away from the body representing referent 1 to represent the dynamic action of referent 2 (Fig. 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 (Fig. 16b). However, even when the dynamic action of referent 2 was used in

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

	Action encoded by hand					
	Signers			Silent gesturers		
Type of action	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%

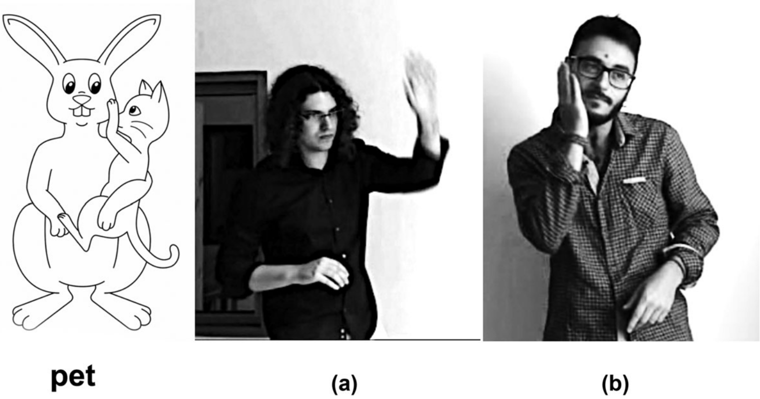

Fig. 15. Strategies used by silent gesturers to encode dynamic action 2, performed by the hand of referent 2 (a: prototypical encoding; b: nonprototypical encoding).

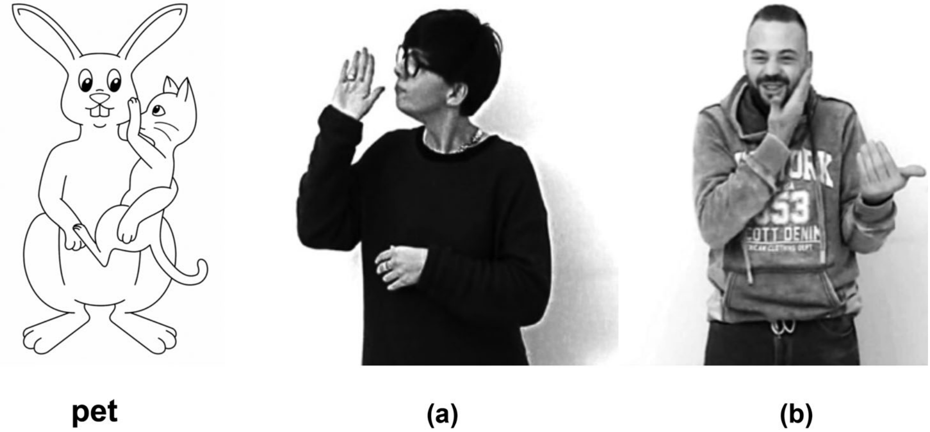

Fig. 16. Prototypical strategies used by signers to encode dynamic action 2, performed by the hand of referent 2.

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 nonmanual articulators, thus increasing the number of simultaneously encoded information units in a single movement segment.

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 (Fig. 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

Fig. 17. Prototypical encoding sequence for *Information density level 4* from gesturers (Bird–Bunny pair).

Fig. 18. Prototypical encoding sequence for *Information density level 4* from signers (Bird–Bunny pair).

consecutive unit was integrated with the preceding one (Fig. 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. 18 MS4) but also by deictically referring to it by means of pointing (Fig. 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.

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.

5.1. Achieving simultaneity in sign language versus 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 seven movement segments in silent gesturers (of which six movement segments were produced by the same participant) compared to 102 movement segments for signers. Thus, even though the affordances of 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 and 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 show 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 Figs. 13 and 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, for example, in order to represent a kissing action silent gesturers would use their mouth (see Fig. 13) rather than using hand to represent kissing as preferred by signers (as in Fig. 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, for example, 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.

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 receiver 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 toward 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 reused 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."

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, Arbib, & Garrod, 2013, 2014; Kirby, Tamariz, Cornish, & Smith, 2015; Motamedi et al., 2019). Below we review how coreferential 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 language users 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, Levy, & Gibson, 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, Jaeger, & Newport, 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. 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.

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, Özyürek, & Morgan, 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.

Acknowledgments

We wish to thank Giulia Demartis and Chiara Madia for data collection and coding. We also 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.

Funding Information

This project was funded by the DCOMM Grant, the European Union's Horizon 2020 research and innovation program under the Marie Sklodowska-Curie Actions Grant Agreement No. 676063.

Conflict of Interest

The authors declare no conflict of interest.

Open Research Badges

This article has earned Open Data badge. Data are available at https://osf.io/dprmf/?view_only=144de5c291b349f29cbf137ec3f0291a.

References

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

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*, 1–48. https://doi.org/10.18637/jss.v067.i01

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

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

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]. https://www.proquest.com/docview/305338656/abstract/D565BDAC93534A09PQ/1

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). Cascadilla Press.

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

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

Cuxac, C. (2000). Recherches linguistiques de Vincennes. *Recherches linguistiques de Vincennes*, *29*, 55–72. https://doi.org/10.4000/rlv.1198

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

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

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

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

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

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.

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

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

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

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

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

Kirby, S. (2000). Syntax without natural selection: How compositionality emerges from vocabulary in a population of learners. In C. Knight, M. Studdert-Kennedy, & J. Hurford (Eds.), *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

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

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

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

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.

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

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

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.

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

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

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

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., 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

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

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

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

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

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.). John Benjamins Publishing Company. https://halshs.archives-ouvertes.fr/halshs-00347426

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

Senghas, A. (2019). How language learns: Linking universals to acquisition. In Proceedings of the 43rd annual Boston University conference on language development. pp. 1–10.

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). In 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. M. Smith, M. Schouwstra, B. de Boer, & K. Smith (Eds.), *The evolution of language* (pp. 279–288). 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 segmentation and sequencing. In R. Botha & M. Everaert (Eds.), *The evolutionary emergence of language*. 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

Siple, P., & Fischer, S. D. (Eds.). (1991). *Theoretical issues in sign language research, volume 2: Psychology*. University of Chicago Press. https://press.uchicago.edu/ucp/books/book/chicago/T/bo3637114.html

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

Supalla, T. R. (1982). *Structure and acquisition of verbs of motion and location in American sign language by Ted Supalla*. University Microfilms.

Taub, S. F. (2001). *Language from the body: Iconicity and metaphor in American sign language*. Cambridge University Press. https://doi.org/10.1017/CBO9780511509629

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

Vermeerbergen, M., Leeson, L., & Crasborn, O. A. (2007). *Simultaneity in Signed Languages: Form and Function* (Vol. 281). John Benjamins Publishing.

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

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Supplementary material
Supplementary material