4 Homesign as a way-station between co-speech gesture and sign language: the evolution of segmentation and sequencing

Ann Senghas, Asli Özyürek,
and Susan Goldin-Meadow

4.1 Introduction

Information can, in principle, be organized in many ways, but certain patterns recur in language after language. One apparently universal organizing principle is the segmentation and combinatorial sequencing of basic categorical elements (Hockett 1960b, 1987). Where does this practice come from? Segmentation and sequencing is not the only way to bundle information. For example, representations such as maps, paintings, and acted-out imitations of behaviours are structured iconically, that is, they derive their organization wholly from their referents. Patterns in such representations correspond, part-for-part, to patterns in the thing represented. Half of a city map represents half of the city, and the initial moment of acting out a behaviour represents the initial moment of the behaviour. In contrast, the sequenced patterns of language do not imitate the world it represents. There is no part of New York City that corresponds to the word 'York'.

In a series of studies, we have been tracing the steps of spontaneous communication systems as they progress from unanalysed, holistic representations to discrete, sequenced elements. The approach of much recent computational and experimental work has been to simulate the emergence of features of this sort in an artificial language (e.g. Christiansen and Kirby 2003). Such an approach provides fruitful springboards for speculation about language evolution, but must be complemented by data from actual

communities where new language systems have emerged *de novo*. Two types of naturally emergent systems appear promising—*homesigns*, the gestural communication systems that develop in individual households containing a deaf member (Coppola and Newport 2005; Goldin-Meadow 2003) and *emergent sign languages*, manual language systems that arise when homesign gestural systems are transmitted between individuals within a generation and across different generations (e.g. Nicaraguan Sign Language, Senghas et al. 2004; Al Sayyid Bedouin Sign Language, Sandler et al. 2005). These naturally developing systems provide unprecedented opportunities to track empirically the steps of human language emergence. We briefly describe characteristics of homesign gesture systems in general, and then one particular emergent sign language, Nicaraguan Sign Language. We focus on how these emerging systems express motion events, a domain that presents rich possibilities for both holistic and segmented representational formats.

4.1.1 *Homesign*

Deaf children born to deaf parents and exposed to a mature sign language from birth learn that language as naturally as hearing children learn the spoken language to which they are exposed (Lillo-Martin 1999; Newport and Meier 1985). Children who lack the ability to hear thus have no difficulty learning language and will exercise their language-learning skills if exposed to usable linguistic input. However, most deaf children are not born to deaf parents who can provide them with a model of a conventional sign language. Rather, they are born to hearing parents, who are unlikely to know a sign language. Children whose hearing loss is severe are typically unable to learn the spoken language that their parents use with them, even when given hearing aids and intensive instruction. If, in addition, their hearing parents do not put them in an educational situation where they will be exposed to sign language, they will have no usable input from any conventional language.

What happens in such a situation? Deaf children around the globe use their hands to communicate with the hearing individuals they know (e.g. Goldin-Meadow and Mylander 1998). These gestures—called 'homesign'—have many of the properties found in natural languages. For example, homesigns have a stable lexicon (Goldin-Meadow et al. 1994), word-level compositional (Goldin-Meadow et al. 1995, 2007) and

morpho-phonological (Brentari et al. in press) structure, sentence-level ordering and deletion rules (Goldin-Meadow and Feldman 1977; Feldman et al. 1978), recursion (Goldin-Meadow 1982; Goldin-Meadow 2005), grammatical (noun, verb, Goldin-Meadow et al. 1994) and syntactic (subject, Coppola and Newport 2005) categories, and negative and question operations (Franklin et al. 2011). Moreover, homesigns are used for many of the functions that conventional languages typically serve—communicating about displaced events (Morford and Goldin-Meadow 1997), telling culturally appropriate stories (Phillips et al. 2001), and making generic statements (Goldin-Meadow et al. 2005), to name a few.

Homesign systems arise when a deaf child is unable to acquire spoken language and is not exposed to sign language. Homesign systems are not shared in the way that conventional communication systems are shared. The deaf child produces gestures to communicate with hearing individuals in the home, but those individuals, particularly in Western cultures, are often committed to teaching the child to talk and use speech whenever they communicate with the child. Although the hearing speakers do gesture when they talk, those gestures form an integrated system with the speech they accompany and thus are not free to take on the properties of the deaf child's homesigns (Goldin-Meadow et al. 1996). As a result, although hearing speakers respond to the deaf child's gestures with gestures of their own, they do not adopt the deaf children's gesture systems. It is in this sense that homesign differs from conventional sign languages, and even from an emerging sign language such as Nicaraguan Sign Language, whose users not only produce, but also receive the signs of their language.

4.1.2 *Nicaraguan Sign Language: an emergent sign language*

In the late 1970s and early 1980s, rapidly expanding special education programmes in Nicaragua brought many deaf children together for the first time (Kegl and Iwata 1989; A. Senghas 1995). Previously, most deaf individuals were isolated in their homes, and the few schools and clinics available served small numbers of deaf youths for short periods, without leading to contact outside school hours (Polich 1998; R. J. Senghas 1997). Consequently, deaf Nicaraguan children had minimal contact with each other, and no contact with deaf individuals older than themselves. In this context, no sign language emerged, evidenced by the lack of a shared language in deaf Nicaraguan adults over the age of 45, even today.

At the onset, in 1977, approximately fifty deaf children were enrolled in the new programmes. This number increased to over four hundred by the mid-1980s (Polich 2005). Although language instruction concentrated on teaching students to lip-read and to speak Spanish (with minimal success), the children spontaneously began to use gestures to communicate with each other. As they interacted socially on school buses, in the schoolyard, and later in their homes, the students converged on a common vocabulary of signs and characteristic ways to express them—and a new language, Nicaraguan Sign Language (NSL), was born. The language has continued to develop and change as new waves of children enter the community each year and learn to sign from older peers. Today there are approximately 1,200 signers of NSL, ranging from 1 to 45 years of age.

This is not an unusual history for a sign language. Other languages have originated in a school context, and been passed from student to student ever since. What is special about the Nicaraguan case is that it occurred recently enough for the originators of the language still to be alive. Taken together with the generation that followed them, they provide a living historical record of a language as it develops through its earliest stages.

For experimental purposes, it has been convenient to divide the community into age cohorts based on year of arrival in the signing community. We define the first cohort as those who arrived in the late 1970s and early 1980s; the second, those who arrived in the mid to late 1980s; and the third, those who arrived since 1990. We will take advantage of this sequence of cohorts to explore the nature of the processes that shaped the language as it was passed from one cohort to the next. In addition, we will examine the gestures produced by Spanish-speakers surrounding the deaf Nicaraguan community as we consider whether they represent a source for the emergent patterns found across the three age cohorts of NSL signers.

4.2 Motion event expressions in an emerging sign language

The description of motion events—such as the way a linguistic expression describes the event of an object rolling down an incline—offers a promising domain for detecting the segmentation and sequencing of basic

elements. Perceptually, rolling is experienced as a holistic, unsegmented event that simultaneously includes rotation and linear displacement. However, cross-linguistic work has shown that languages typically separate expressions of complex motion into elements that encode the manner and the path of motion, and combine these elements according to the rules of the particular language (Talmy 1985). For example, English produces one word to express manner (*rolling*) and another to express path (*down*), and assembles them into the sequence *rolling down*. In an initial study (Senghas et al. 2004), we examined whether NSL, over the course of its early development, represented motion in a holistic, iconic manner faithful to the physical motion, or with the discreteness and combinatorial structure typical of developed languages.

We asked ten signers from each of the first three cohorts of NSL to describe a collection of videotaped motion events, such as a cat climbing up a drainpipe or rolling down a hill. We also asked ten hearing Nicaraguans to describe the same events in Spanish, and observed their co-speech gestures. Each participant watched an animated video cartoon that included these events, and narrated its story to a peer. Deaf participants signed their narratives. Hearing participants spoke Spanish, and only their co-speech gestures were analysed. For all of the narratives, the expressions that described the motion events were analysed with respect to how the different aspects of the motion were included. Specifically, we determined whether information about manner and path was (A) a conflated, simultaneous expression, with a single hand movement, or (B) a sequence of manner-only and path-only elements. Examples of these two types of expression are shown in Figure 4.1. Note that a single response could include both types.

We found that Spanish-speakers always, and first-cohort signers often, produced manner and path together as a single holistic, conflated gesture movement. In this way, their expressions matched the structure of events in the world. In contrast, second- and third-cohort NSL signers preferred to separate events into sequences of pure, elemental manner-only and path-only signs (see Figure 4.2). These new expressions already include the segmentation and sequencing characteristic of language.

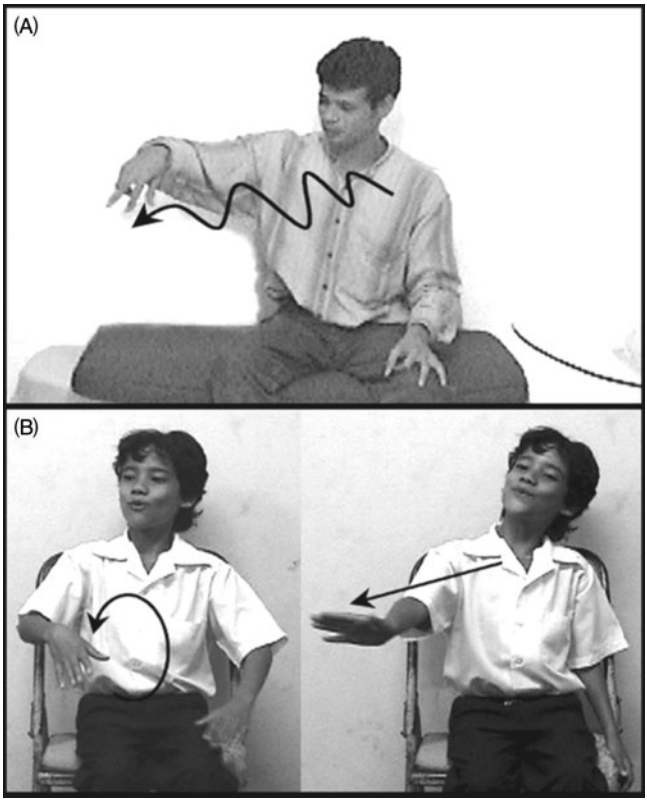


FIG. 4.1. Examples of motion event expressions from participants' narratives. (A) Manner and path expressed simultaneously. In this example a Spanish-speaker describes a character rolling down a hill with a bowling ball in his belly; the gesture shown was produced along with speech. Here manner (wiggling) and path (trajectory to the speaker's right) are expressed together in a single holistic movement. (B) Manner and path expressed sequentially. In this example, a third-cohort signer describes the same rolling event in NSL. Here manner (circling) and path (trajectory to the signer's right) are expressed in two separate signs, assembled into a sequence (from Senghas et al. 2004).

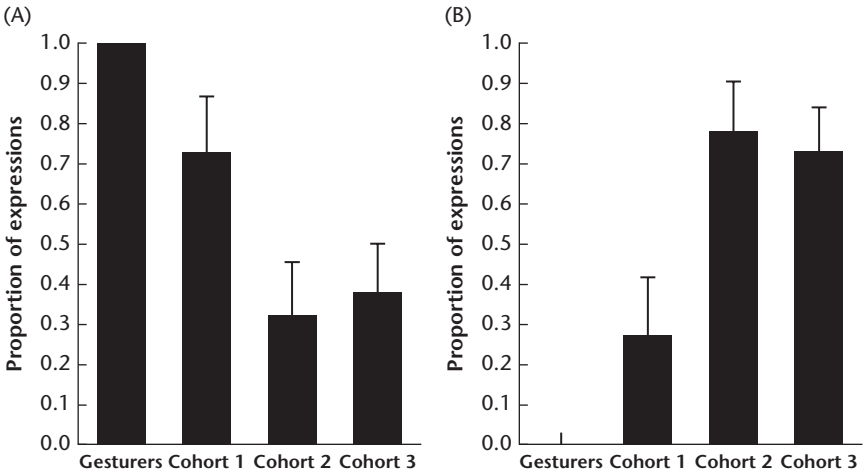


FIG. 4.2. Conflated and sequential expression of manner and path.
(A) The proportion of expressions with manner and path in which the two are conflated within a single gesture or sign. Bars indicate mean proportions for individuals in each of the four groups; error bars indicate SE. All of the co-speech gestures and most of the first-cohort signers' expressions conflate manner and path. Second- and third-cohort signers produce relatively fewer expressions of this type. (B) The proportion of expressions with manner and path in which the two are produced sequentially as manner-only and path-only elements. Such sequences are never observed in the co-speech gestures. First-cohort signers sometimes produce them; second- and third-cohort signers include them in most of their expressions (from Senghas et al. 2004).

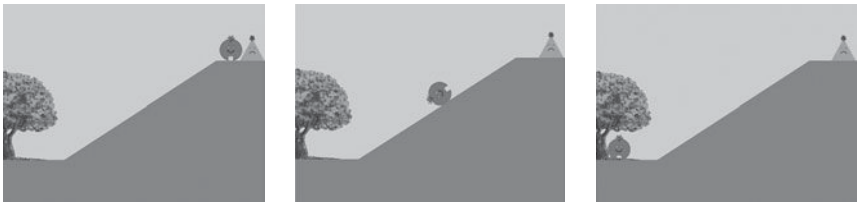
4.3 The missing step: motion event expressions in homesign

The holistic manner and path conflation produced by the Spanish-speakers in the gestures that accompanied their speech contrasts starkly with the segmented manner–path sequence produced by the second- and third-cohort Nicaraguan signers. Where did the segmented and sequenced patterns characteristic of NSL come from? Assuming that co-speech gestures of hearing people were one source of input to the emerging sign language used by deaf people, we are faced with an abrupt transition between the gestures of Nicaraguan speakers and the signs of Nicaraguan signers. There is, however, a transitional step. Almost certainly, prior to coming together for the first time in 1977, the Nicaraguan deaf children

had been using gestures to communicate with the hearing people in their households—they were homesigners. The question we ask here is whether the deaf individuals who came together to form the first cohort of NSL had already begun, as homesigners, the process of segmentation that has come to characterize NSL.

We have not yet explored this possibility in Nicaragua today. However, we have studied comparable homesigners in Turkey. We identified seven Turkish homesigners, ranging in age from 3;2 to 5;6 (years;months), who had learned neither a spoken nor a signed language. These homesigners were shown short animated video clips of motion events highlighting manner and path (see an example of a target event in the middle panel of Figure 4.3, from Özyürek et al. 2008) and were asked to describe what happened in each clip. During their narration, the children were given a picture of the initial scene of each event so that, if necessary, they could use pointing gestures to refer to the characters in the event. Children were videotaped at home every one to three months. The descriptions analysed for this study come from six sessions for each child, conducted over the course of a year. All of the children were congenitally deaf, with bilateral hearing losses (70–90 dB) and no other reported cognitive or physical disabilities. The children's hearing parents had chosen to educate them using oral (i.e. non-signing) methods. At the time of our study, the children had received minimal or no speech therapy and, although they were able to produce an occasional Turkish word, did not combine words into sentences. In addition, none had been exposed to conventional sign language or had contact with another deaf child or adult.

We coded all of the gestures that the children used to convey motion information, and classified each gesture into one of three types: (i) manner gestures, e.g. the hand rotates in place; (ii) path gestures, e.g. the hand moves across space in a straight path; (iii) manner + path (conflated) gestures, e.g. the hand rotates while moving across space in a straight path. In many cases, the children enacted the manner or path of motion, or traced it on the picture that they were given; such responses were also coded as manner and/or path gestures. Using criteria developed by Goldin-Meadow and Mylander (1984), we divided the gestures into sentence strings, and classified each sentence that contained information about both manner and path into one of three types, based on how that information was combined: (i) Conflated only (containing only manner + path gestures); (ii) Sequenced only (containing both manner



Entry event: **Green Man hits Tomato Man**

Target event: **Tomato Man rolls down hill**

Closing event: **Tomato Man slides and hits tree**

FIG. 4.3. A sample video designed to elicit descriptions of motion events. Note that the target event involves both manner (roll) and path (descend).

gestures and path gestures and no conflated gestures; (iii) Mixed (containing a conflated gesture plus a manner gesture and/or a path gesture).

We found that almost half (49%, SD=18%) of the Turkish homesigners' sentences contained Conflated gestures alone, and relatively few (14%, SD=19%) contained Sequenced gestures alone. In this sense, the homesigners' pattern resembled the pattern found in the first cohort of Nicaraguan signers (cf. Figure 4.2). However, a sizeable percentage (34%, SD=15%) of the Turkish homesigners' gesture sentences were of the third, Mixed type—a conflated gesture combined with one or more elemental components (see Goldin-Meadow et al. under review, for evidence that Turkish and American homesigners produce the Mixed form in spontaneous communication). Thus the homesigners appear to be in a transitional period with respect to segmentation—they were able to segment an action component out of the conflated motion, but they also produced the conflated form along with the segmented form (see Figure 4.4A).

If the gestures that the Turkish homesigners produce—their Mixed gesture expressions, in particular—reflect an early stage in the emergence of a language system, then we might expect to find Mixed expressions also in the first users of Nicaraguan Sign Language, that is, in first-cohort signers. To test this possibility, we reanalysed the Nicaraguan data reported in Senghas et al. (2004), this time classifying the gesturers' and signers' expressions into the same three types—Conflated, Mixed, and Sequenced. The results are presented in Figure 4.5.

The hearing Spanish-speakers relied predominantly on Conflated gestures to convey manner and path information. In contrast, the second- and third-cohort Nicaraguan signers relied predominantly on Sequenced

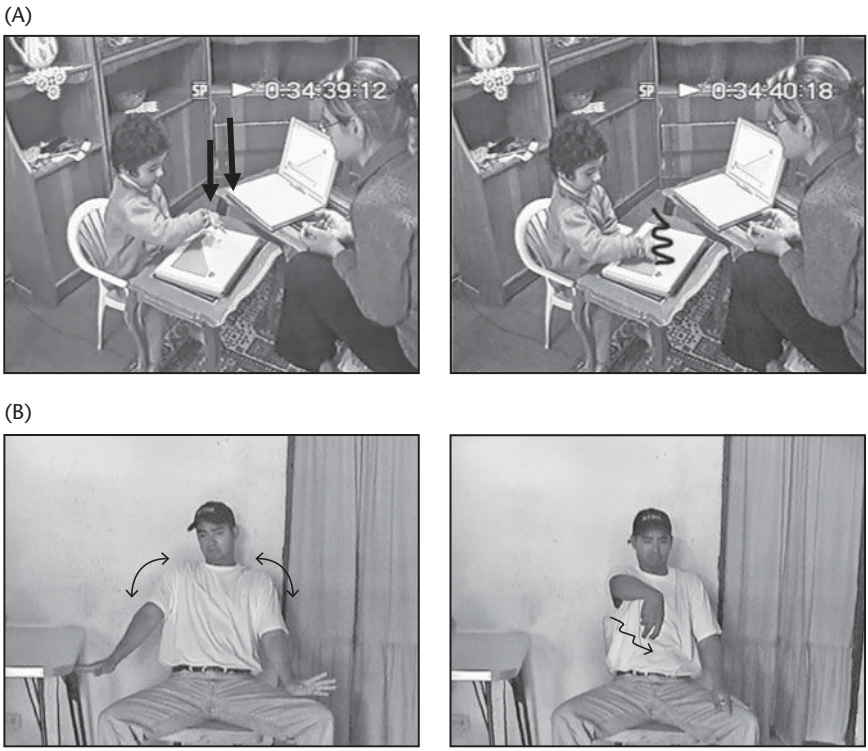


FIG. 4.4. Examples of Mixed gesture sentences.
(A) A Turkish homesigner describes the jumping-up movement of a triangle figure. She first produces a gesture for the jumping manner, followed by a gesture conflating the jumping manner and the upward path. (B) A first-cohort NSL signer describes a character rolling down a hill with a bowling ball in his belly. He first produces a body gesture for the side-to-side waddling manner, followed by a gesture conflating both the waddling manner and the forward path.

gestures to convey manner and path information. The interesting group is the first cohort, who, as predicted, appear to be in a transitional state between the gesturers on the one hand, and the younger signers on the other. The first cohort's preferred way of conveying manner and path information is to produce Mixed gesture expressions, that is, sentences containing a conflated manner + path gesture along with a segmented manner or path gesture (see Figure 4.4B).

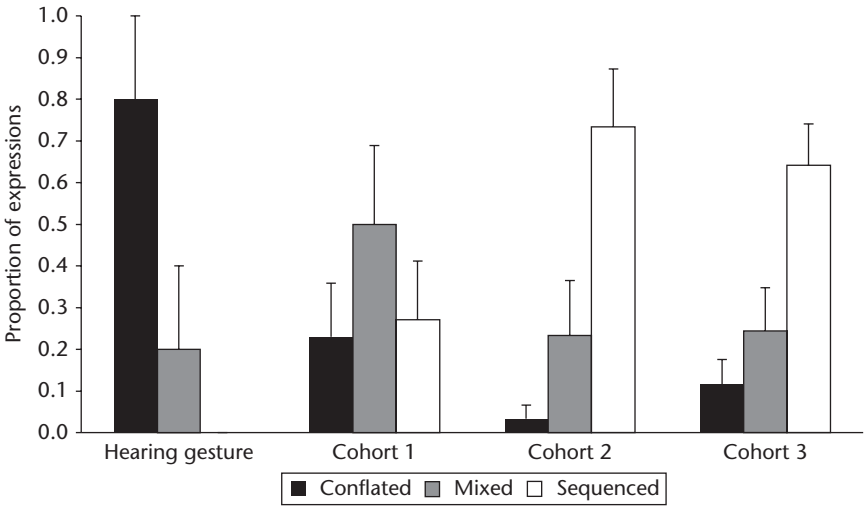


FIG. 4.5. The proportion of expressions produced by the hearing Spanish speakers and the three cohorts of Nicaraguan signers, classified according to the segmentation of the manner and path components: Conflated (no segmentation), Mixed (partial segmentation), Sequenced (full segmentation). Error bars indicate SE.

4.4 Discussion

By studying natural, present-day, emergent language systems, we have been able to capture the earliest stages in the development of one fundamental property of human language: segmentation and sequencing. Homesigners presented with holistic, conflated gestures do not faithfully reproduce the gestures they observe in their environment. Instead, they segment those holistic gestures into components, and combine the components into a new kind of structured utterance. The roots of this process can be found even in a single-member communication system, that is, in the first step taken by individual homesigners.[1]

We begin by assuming that the co-speech gestures the homesigning child sees provide input to the child's own signs. We suggest, however, that

[1] Note that we are making no claims about segmentation at the level of the signal—what is typically called phonology (i.e. Hockett's 1960b duality of patterning feature). Our claim is at the level of semantic components (see Sandler et al. 2011 for evidence that the emergence of phonology in another developing sign language, ABSL, is slow and gradual).

the homesigning child does not veridically copy this input. Although our data on co-speech gesture come from Spanish-speakers and our data on homesign come from Turkish children, the findings are nevertheless suggestive. The co-speech gestures contain many instances of the conflated form and only a few instances of the mixed form (Figure 4.5). In contrast, the mixed form appears relatively frequently in the homesigners' expressions, and even instances of the sequenced form are evident (recall that these did not occur at all in the Spanish speakers' co-speech gestures, Figure 4.2B). In future work, we will examine homesigners in Nicaragua in order to explore the relation between gesture when it is used along with speech (in hearing speakers), and gesture when it first begins to take over the full burden of communication (in homesigners). We can thus explore which forms emerge and become prevalent as gesture takes on the full functions of language.

Humans can certainly learn to understand and produce analogue representations (e.g. pictures and maps) and can even integrate those representations with linguistic representations (e.g. co-speech gesture). Nevertheless, deaf homesigners seem to transform the holistic analogue representations from co-speech gesture into segmented and sequenced forms. Importantly, this transformation is not an inevitable response to stimuli of this sort; otherwise co-speech gesture and other analogue representations would all have taken on a segmented form generations ago. Instead, the new mixed gesture form, which is a combination of the analogue and the segmented, takes hold only when gesture assumes a different function—when it assumes the full burden of communication and becomes a primary language system (Goldin-Meadow et al. 1996).

Following this first step in language emergence, we suggest that two patterns of transmission are needed for the gesture system to continue to develop and converge on a more mature, segmented language-like form. The first is horizontal transmission across peers within a single generation. The effect of this process is evident in the contrast between homesigners and first-generation NSL signers—first-cohort Nicaraguan signers segment somewhat more than Turkish homesigners. If this difference holds true when we observe Nicaraguan homesigners, we can speculate that reciprocal interaction among members of the community (e.g. being both a producer and receiver of the communication) favours the emergence of language-like forms—in this case, a progression from holistic to segmented and sequenced signs. This progression was recently

documented in an experimental study of the process by which iconic graphical signs become symbolic signs. Garrod et al. (2007) found that the degree of interaction (i.e. feedback among participants), rather than mere repeated usage with no interactive partner, facilitated the emergence of symbolic signs and the loss of iconicity. We speculate that the number of interactive partners in the first cohort in Nicaragua (which is far greater than the number in a homesigning situation) drives the emergence of signs that are more conventionalized and categorical (and thus less iconic)—specifically, signs that are segmented and depict elements of manner and path rather than a holistic image of the motion event.

The second pattern needed for continued language emergence is vertical transmission from one generation to the next as new learners enter the community. The effect of this process is evident in the differences in the preferred forms across the three cohorts of NSL signers—a steady increase in segmentation from partial (mixed) to full segmentation. It is relevant that all of the learners of NSL in this study, including the first-cohort signers, were children at the time of learning; that is, they converged on a system with peers as children, and subsequently passed the system on as adolescents and adults to new children. Evidently, child learners have a natural inclination to analyse a linguistic signal as discrete and combinatorial, even if it is originally presented as continuous and holistic. Crucial to this inference is recognizing that if adults were as likely as children to apply this analysis, or if the analysis were a consequence of merely interacting with other members of a community, we would see segmentation used equally often across all the cohorts in Figure 4.5. We did not. The pattern seen in the figure instead points to a language-specific learning strategy that is particularly available early in life.

Christiansen and Chater (2008) propose that characteristics universally observed in languages, such as the segmentation we observed here, reflect characteristics of the human mind. On this view, languages develop by adapting to the nature of language learners. As a corollary, they suggest that there is no need for special learning devices to have evolved that specifically apply to language learning. Compositionality, they argue, is a product of the combinatorial structure of human thought, and sequential ordering is the product of the 'seriality of vocal output'. If this view is adapted to historical language evolution, the only change that took place was the adaptation of the signal itself to general human processing abilities.

The data we present here enrich, and belie, this account. The segmentation and recombination processes we observe are not inevitably applied to incoming gesture—the transformation occurs only when the signal is taken to function as a primary communication system, that is, as language. The medium of gesture can easily handle a conflated, simultaneous representation (conflation is, in fact, typical of Nicaraguan co-speech gesture). Consequently, there appears to be no pressure for this input to be transformed into a longer, segmented form. Moreover, the qualitative difference between child and adult learners exposed to NSL points to a process that is not necessarily a general cognitive one. Thus, if we extrapolate from the language emergence patterns we see in our data to historical language evolution, we are led to hypothesize that learners and languages co-evolved, resulting in a human mind that is particularly adept at learning the kinds of languages humans produce. One product of this process is an analytical approach to any input that serves as a language, a process that breaks the input down piece by piece, eventually arriving at fundamental elements.

We do not claim that the process of language emergence we see today in modern systems is necessarily a re-enactment of the original process of evolution of language (e.g. Arbib 2005), though ultimately both processes, on very different time scales, are likely to lead to the segmented product we see today. It seems unlikely that a segmented and sequenced format evolved in the original first language as quickly as it has in these modern emerging sign systems. Rather, we speculate that over an extended period of co-evolution learners and languages both came to favour an analytical, combinatorial pattern. Once languages began to take on this form, children with a bias to analyse and segment language in this way would have a learning advantage, and these learning mechanisms would then be favoured over time. As a consequence, once the bias is in place, any new languages to emerge would quickly take on a segmented, combinatorial structure, making it a universal linguistic feature. In modern humans, this analytical approach is applied in every instance of language learning or language creation, enabling learners to extract basic elements from a stream of input. Segmented structures will, as a result, dominate after a few short generations of transmission, as we see in the newly emerging systems described here. This is the imprint of the human mind on language.

Acknowledgements

The research described in this chapter was supported by grants from the National Institutes of Health (NIDCD) R01-DC05407 to Ann Senghas and R01-DC000491 to Susan Goldin-Meadow. The content is the responsibility of the authors and does not necessarily reflect the views of the funding institutions. We are grateful to Burcu Sancer and Reyhan Furman for help in collecting and coding the Turkish homesign data; Koc University, Istanbul, Turkey, for logistical support; and the participants and their families, for sharing their time and their language with us.