Published in final edited form as:

Can J Linguist. 2011 March 1; 56(1): 109–124.

The problem of conventionality in the development of creole morphological systems

Carla L. Hudson Kam and
University of British Columbia

Whitney Goodrich Smith
University of Maryland

Abstract

This paper focuses on the problem of the conventionalization of grammatical morphology during language formation, asking how a form comes to have a shared grammatical meaning in an emerging linguistic community. Results from an experiment, inspired by changes in the use of space in Nicaraguan Sign Language, are presented that demonstrate the difficulty of conventionalization for grammatical forms. We show that even in cases where meaning should be relatively easy to discern, adult listeners are not consistent, either internally or when compared to each other, in the inferences they draw. The paper ends with a discussion of the nature of the problem with respect to emerging languages more broadly, and speculates on how grammatical forms might come to have shared meanings in a newly emerging language, focussing on the role that children might play in the process.

1. Introduction

Although long-standing belief held that creole languages lack productive morphology, recent work shows this to be untrue. Indeed, thanks to very careful investigations by numerous researchers, it is now quite clear creole languages have both derivational and inflectional morphology (see, e.g., DeGraff 2001; Lefebvre 2003; Plag, 2003, 2006). However, it is also clear that the creole forms are generally not the same as in the lexifier languages, in terms of their form, meaning, conditions of use, or some combination of these, which raises interesting questions with regard to the development of these languages, the answers to which are important not just for our understanding of these languages, but also for our understanding of the human capacity for language learning and creation more broadly. Questions related to the nature and source of the morphological forms are of natural interest to linguists, for instance, did the morphemes arise through borrowing or grammaticalization within the creole? Are creole morphemes more semantically transparent than those in non-creole languages? (See Plag 2006 for some discussion of these themes.)

Other issues remain, however, even after these questions have been answered; for instance, how does a form that is likely variable in its use becomes a stable part of a language's grammar, shared by all speakers, or why that form and not another (particularly when a substrate form is borrowed)? While these latter questions might seem to be psycholinguistic or even social-psychological rather than strictly linguistic questions, understanding these other processes is also necessary if we are to truly understand the full dynamics of language emergence, formation, and change (Mufwene 2010).

Here we explore one such process: how a form comes to have a shared meaning in an emerging linguistic community – that is, the emergence of conventionality (Hockett 1960). Our discussion is inspired by the development of spatial morphology in Nicaraguan Sign Language (NSL); however, this is an issue that is relevant for grammatical forms in any new contact language. Indeed, one might assume that it is less of a problem in a signed language where some degree of iconicity can get around the problem of discerning meaning for novel forms, since the meaning of signs (iconic ones at least) is often quite obviously related to their form.¹ Alas, as we will show, even here it remains a problem.

We begin by describing the nature of spatial morphology in NSL, a signed language that developed in Nicaragua in the 1980s. Although not a typical creole, in that it lacks a lexifier or superstrate language and substrates shared by multiple participants involved in the language creation process (Kegl and McWhorter 1997), NSL has nevertheless been described as a creole by some researchers (see, e.g., Senghas 1995). Whatever its status, NSL presents an interesting case for investigating the emergence of conventionality in a morphological system. After briefly describing how space is used in the language at present, we discuss why it is surprising (at least to us) that the forms in question took time to emerge and/or stabilise in NSL, and present results from an experiment with adult hearing speakers exploring why this might be the case. We end by discussing the nature of the problem of conventionalization with respect to emerging languages more broadly, and speculate on how grammatical forms might come to have shared meanings in a newly emerging language.

1.1 Background on Nicaraguan Sign Language

The initial social conditions enabling the emergence of Nicaraguan Sign Language occurred in the late 1970s when an elementary school for deaf children was established. Although the school was oralist, the children were allowed to gesture with each other in the playground. Many of these same children continued on to a vocational school that was established in 1980, and importantly, began to socialize outside of school hours, using the emerging gestural system that they were actively involved in creating. The community, and with it the language, continued to grow as time went on and more and more children and adolescents had the opportunity to attend the two schools. Eventually, a Deaf Association was formed, providing a locus for the Deaf community post-school (see Senghas et al. 2005, for a more complete description of the history of the language).²

Like many signed languages (Padden 1983; Supalla 1982), NSL now includes the use of space in its grammar (Senghas 2003). The way in which multiple entities are situated in signed space corresponds in a meaningful way to their actual locations (relative to each other and/or the signer) (Senghas and Coppola 2001). In particular, signers use a rotated perspective, and listeners (or more correctly in this case, viewers), interpret the sentence accordingly. An example from Senghas (2003: 516–518): signers watch a video in which a woman gives a cup to a man seated to her right a cup. Note that the man is to the signers' left as they view the video. When asked to describe the event, signers produce 'pass the cup' to their right, not their left. In a second step, listeners asked to interpret the description will understand it as referring to a woman passing a cup to her right, not their own right. However, this is true only for signers who learned NSL as children and were part of the second cohort of language users, so after 1983.³ Signers who learned the language prior to that time or as adolescents are less consistent in how they sign space, both internally and

¹Although many signs are iconic, they are still arbitrary conventions in the sense that that sign, and only that sign, will do; signs are not pantomimes. Moreover, signs often lose their iconicity over time, either as the signs change, or the cultural practices that led to them change (e.g. the sign for *girl* in ASL is a trace of a bonnet string, something no longer part of a girl's typical dress).

²Following established convention in the literature, deaf with a lower case <d> is used to refer to the medical condition of hearing loss, in contrast to Deaf (upper case D), which connotes cultural affiliation (see Senghas and Monaghan 2002).

when compared with other signers, and fail to interpret spatial modulation in signs (Senghas 2003). It is as if spatial modulations are merely movement that is an inevitable part of the act of signing for them.⁴

This has important implications for other aspects of the language. When space maps onto meaning in a consistent way it can be used to link up other predicates with the same entities; for example, a signer may now perform a sign with respect to the space previously associated with the man (as recipient of the cup) to indicate that he performed or was the patient or recipient of some other action (Senghas and Coppola 2001). Note that whether the man is located to the left or the right (of the other participants or the signer) is irrelevant; what is important is that the same space is used. Crucially, this only works as a grammatical device if everyone involved can both remember what space was associated with which entities and can understand space as it corresponds to viewpoint in the same way. Not surprisingly, then, this coreferential use of space is apparent only in those who learned the language after 1983 and as children (Senghas and Coppola 2001). This grammatical use of space is not a necessary consequence of consistency in the use of space and perspective, but consistency in the use of space and perspective is a necessary precondition of the former.

1.2 A Spatial Puzzle

Several facts about the use of space elsewhere make its developmental timeline in NSL somewhat puzzling. Space is used grammatically in almost every signed language studied to date (see Newport and Supalla 2000). It is also used by some homesigners (home-sign is a gestural communication system created by individual deaf children in hearing environments who are not exposed to a sign language, Goldin-Meadow 2003b). This suggests that it is a very natural aspect of signed grammar (Coppola 2002; Coppola and So 2005; Goldin-Meadow 2005). Moreover, the use of space is not limited to signing; hearing people with no experience with signed languages use space co-referentially when gesturing both with (So et al. 2009), and without (So et al. 2005) speech, and spatial information in co-speech gesture is interpreted co-referentially by hearing listeners with no exposure to signed languages (Goodrich and Hudson Kam 2007; Goodrich Smith and Hudson Kam 2010). Taken together, these findings suggest that the use of space, both in production and comprehension, is part of the general human cognitive endowment, although potentially specific to communication. Why then was it not a part of NSL grammar from the very beginning? We suggest that the delay is due to issues of conventionalization, that it took some time for the meaning of space to become shared by users of the language.

In order for space to be used in the grammar of a language, several conditions must be met. It must be interpreted, consistently, and in the same way, by all users. That is, the comprehender must have the ability to link entities with a spatial location, remember which entities are located where in the abstract signing space, and then recognize those locations across clauses or utterances. There is evidence for this first precondition in some of our previous work (Goodrich and Hudson Kam, 2007, 2010; Goodrich, 2009), where we found that co-speech gesture can affect the interpretation of otherwise ambiguous pronouns by people not exposed to a signed language (and so who have not learned to attend to spatial gestures). This seems, then, to be a basic human cognitive ability, not one dependent on language or input. But this ability is not enough. The comprehender must also interpret the

³According to Senghas (2003), it was around this time that the first intergenerational contact occurred, and with it, the language was passed along to a new cohort of children. It is thus frequently used as the dividing line between the first and second cohorts of NSL users.

⁴The picture is somewhat more complicated if we consider descriptions of scenes without human actors. Here even the second cohort is not consistent across signers, although individuals are fairly internally consistent. First cohort signers are consistent neither individually nor across signers (Pyers et al. 2010).

viewpoint consistently, or in the same way for every distinct conversation. Furthermore, every user of the language must do so in the same way.⁵ In the present study we investigate whether these latter two conditions are also generally met. In particular, we investigate whether hearing individuals with no exposure to a signed language interpret spatially modulated gestures consistently, and in the same way as other individuals in the same community.⁶ If they do, it would suggest that these aspects of spatial mapping are also part of the broad human cognitive endowment. If not, then it is consistent with our contention that conventionalization was required even in the case of space in NSL, and, importantly, that we can learn something about conventionalization more broadly from this case.

We had adult native-English speakers watch a narrator tell several short stories, each involving two characters. As a character was introduced, the narrator gestured towards a distinct (within the narration) location in space, either to her left or to her right. This same spatial location was gestured back to whenever that character was mentioned again. Thus, the different characters occupied distinct locations in space. We then asked participants to select the picture that best represented the story that they had just heard. The pictures always contained the two characters, but varied in the left-right orientation of the depiction of the scene or action.

We were primarily interested in whether listeners would consistently choose pictures that maintained the narrator's viewpoint or their own. To illustrate, if the narrator's gesture is to her right when mentioning the girl (the listener's left, as seen on the video monitor) would listeners select the picture with the girl on their left (listener perspective), or with the girl on their right (speaker perspective). Note that neither perspective is necessarily right or wrong. It simply provides a way to code the data, allowing us to observe whether each listener is consistent in the perspective taken, and if so, if different listeners are all consistent in the same way, that is, take the same perspective.

2. Method

The research methodology of the present study is described in this section.

2.1 Participants

Twenty undergraduate students from the University of California, Berkeley took part in this experiment. All were native English speakers (defined as having learned English before age five), although 5 were bilingual (Cantonese (2), Japanese, Mandarin and Vietnamese), and 7 others had at least some experience studying another language. Crucially, none of the participants had studied any signed languages.⁷ Participants received course credit for participating.

2.2 Materials and Procedure

Participants were seated in front of a 21" monitor, and given the following instructions: "You are going to watch a video of a narrator telling 12 short stories. After each story the video will pause, and you will see two pictures labelled A and B. Your job is to select the

⁵It is not the case that all sentences in a signed language have the same perspective. However, perspective shifts are part of the language, that is, they are marked in a consistent way, and so a comprehender knows how to interpret the spatial modulation (see Pyers and Senghas 2007).

⁶The facts about the development of NSL would suggest that these conditions are not generally met; however, it is possible that the earliest users of NSL learned to ignore spatial modulation. They faced quite a task, developing the language, most lexical items and all, de novo. This could have created a situation where space was attended to only insofar as it was important for lexical distinctions, e.g. the orientation of handshapes, resulting in its being ignored for anything else. Alternatively, they could have learned to ignore it because of the variation present in the perspective presented by other signers.

⁷We avoided people with exposure to a signed language since that exposure might lead them to focus on spatial modulation.

picture that you think best represents the story you just heard. I'll be writing down the picture you chose. There are no right or wrong answers, we're just curious to see how you interpret the stories."

Each narration lasted approximately 6 seconds and described two characters each performing a different action. We ensured that the two characters were easily discriminable, for instance, by being different genders or commonly known in popular culture. For instance, one of the narrations presented was, "Bobby gave Andrea a new CD for her birthday. When they listen to the CD, Bobby sings and Andrea dances."

The twelve narrations each had a slightly different structure in order to avoid participants getting bored by the stories. Six were about humans, two were about animals (e.g., puppy and kitty), and four were about well-known (superhero or animal) characters from movies and TV. Narrations contained from two to five sentences (mean = 2.83) (it depended to a large extent on the length of the sentences, which tended to be longer in narrations with fewer sentences). Typically, each character was mentioned by name two to three times in the narrative (mean = 2.16). Although it may seem somewhat odd to use a name multiple times, we did so to make it quite clear who was doing what.

The narrator was shown from the waist up with her hands and face clearly visible. While saying each character's name she produced a localizing gesture, situating the two characters in space with opposite hands. She later gestured back to those same locations when mentioning the characters(s) again. Example gestures from the narration described previously are shown in Figure 1, and a transcript of the actual narration with accompanying gestures is presented in example (1). The period of speech during which the gesture occurred is marked by the square brackets, and the gesture that was produced is described underneath the relevant word(s), enclosed in square brackets. RH stands for "right hand" and LH for "left hand".

(1) [Bobby] bought [Andrea] a new CD for her birthday.

RH palm up

LH palm up

When they listen to the CD, [Bobby sings], and [Andrea dances].

RH palm up

LH palm up

The number of distinct gestures associated with a character varied from narrative to narrative, ranging from three (two for one character and one for the other) to seven (four associated with one character, three with the other).⁸ The mean number of gestures per narrative was 4.33.⁹ Localizing-type gestures such as the ones we used in our stimuli do occur in natural speech contexts (So et al. 2009); however, we chose to create our own video stimuli rather than collect naturalistic examples to ensure that the gestures were as similar as possible across stories (and versions, as described below). Although the gestures in our stimuli were more characteristic of storytelling than conversation (they were slightly more exaggerated than those typical of conversation), they were quite typical for the narrator (the second author).¹⁰ We did not intentionally control the nature of the gestures used – the narrator produced the gestures that were most natural for her given the story – other than to

⁸Both the three and seven gesture narratives contained three sentences.

⁹Occasionally, a gesture was held for a longer period of time and so the same gesture was 'used' in two distinct sentences or clauses, as in the following (the period of speech during which the gesture occurred is marked by the square brackets and the gesture that was produced is described inside the parentheses): My cat's name is [Fluffy. Fluffy has grey fur] (RH palm down, pulses slightly with each mention of name). These were counted as one single gesture. Such held gestures are not uncommon in natural speech contexts, where they appear to maintain a character in the conversational frame of reference, among other things (see Park-Doob 2010).

¹⁰The exaggerated nature of the gestures actually makes them more like the kinds of movements used by signed languages, and so the kinds of movements NSL is developing.

ensure that they were the same in the both versions of the story. Thus, although most of the gestures were of the type shown in Figure 1, produced with an open palm facing upwards, downwards, or inwards and held slightly, five of the 12 narratives also contained gestures with a different hand configuration/movement, in accord with the scene being described, as shown in transcription (2).

(2) [Amy] and [John] are driving bumper cars.

LH palm facing inward

RH palm facing inward

[Amy was surprised] [when John bumped into her] really hard.

LH palm up

RH fist moves quickly across body and 'bump's into LH

There were two counterbalanced versions of the stimuli, which varied only in the hand used with each name. For instance, for the narration described above, in Version 1 the narrator gestured with her left hand while saying the name Bobby and her right hand while saying the name Andrea. In Version 2, the narrator gestured with her right hand for Bobby and her left hand for Andrea. Any co-referential gestures (gesturing back to the same location upon further mentions of the same character) were matched to the initial localizing gesture, and so also varied by version. Ten participants saw one version of the stimuli, and the other 10 saw the other version, such that any individual participant saw only one version of the stimuli. The stimuli were further counterbalanced such that in both versions the first name co-occurred with the left half of the time, and right the rest of the time.

After each narration participants were shown two pictures on the computer screen and asked to indicate by pointing or saying "A" or "B" which of the two pictures best represented the story they had just heard. The pictures represented activities or scenes described in the preceding narrative, and were mirror images of each other. The test pictures used in the narration in (1) are shown in Figure 2. The actual test pictures were each approximately 6" × 4", in color, and centered in the screen, with picture A presented above picture B (not side by side as in the figure, done here in the interests of space). This positioning made it impossible to match the absolute location of any single gesture on the screen to the absolute location of a character in a picture, forcing participants to answer based on the representations they had formed.

After completing all 12 narrations, participants were asked what they thought the study was investigating, allowing us to see if participants noticed anything unusual about the gestures used in the study. Slightly more than half (12/20) did, in fact, comment on the narrator's gestures. Those that did not tended to say something about the pictures being mirror images of each other. There were no differences in responses between those who did and did not comment on the gestures. The entire session lasted approximately 15 minutes.

2.3 Coding

Each participant's selections were coded as taking either the speaker perspective or listener perspective, depending on whether he/she selected the picture that matched the speaker's use of space (which required a rotation in perspective), or his/her own view. For instance, in Version 1 of the example described above, when the narrator said the name Bobby she gestured with her right hand, meaning that the gesture appeared on the left side of the video monitor. Responses in which participants selected the picture with the singing male character on the left side (Fig. 2, Picture A) were coded as listener perspective, and those in which participants selected the male character on the right were coded as speaker perspective (Fig. 2, Picture B). Coding was the opposite for the other version of the stimuli.

3. Results

Figure 3 shows the number of participants who selected the picture corresponding to the speaker perspective 0 – 12 times. This gives a sense of both how internally consistent individuals were, by showing how often they took that viewpoint, and whether they were internally consistent in the same way, by showing how many people selected the viewpoint that number of times. A person who consistently adopted the speaker perspective would score 12/12. Likewise, someone who consistently adopted the listener perspective would score 0/12. Both of these would be consistent response patterns, just from different perspectives. Likewise, people who are between 1 and 11 on the scale are not completely consistent in terms of the perspective adopted in their responses.

We plot the data in terms of speaker perspective because i) this is the perspective that emerged in NSL, and ii) we expected that it would be the most likely interpretation.¹¹ Our intuition was based on the following logic. Imagine seeing a person describing an event involving two animates (and so entities with distinct fronts and backs) in which she herself is not involved. The most likely perspective that the speaker would have on the event is the front, that is, viewing the characters from their fronts or sides, not their backs. Given this, if I, as the listener, understand the event as taking place with the same relative rather than absolute configuration (which seems the most likely for native English speakers, Li and Gleitman 2002, see also Brown and Levinson 1993), I should rotate the actual spatial locations in my mental model of the event. Therefore, if a speaker describes someone as being on her left in an event, as a listener I should form a mental model of the event in which that person is on my left. The opposite interpretation seems much less natural, since it presumably involves thinking of the characters' backs as being towards the speaker, if the speaker's perspective is considered at all. Note that as there are only two possible choices, a plot of listener-perspective responses would look the same, only reversed, e.g., if only one person has 10 speaker-perspective responses, then only one person has two listener-perspective responses.

With respect to the questions we set out to answer, we can observe that no participant was completely consistent in his or her choices; no participant selected pictures according to either speaker or listener perspective for all 12 narratives. It is not the case that responses were random, however; there is a definite trend towards participants not selecting the pictures representing the speaker perspective: Across participants, the mean number of speaker perspective responses was only 3.3 (sd=2.5), which is significantly different from chance responding (6/12), $t_{2\text{-tailed}}(19) = 4.8$, $p < .0001$.¹²

Indeed, half the participants (10/20) selected the speaker-perspective version of the picture only 1 or 2 times out of a possible 12. A further six participants selected the speaker-perspective picture 25%-33% of the time, somewhat less consistent, but still exhibiting the same general pattern of responses. Thus, the majority of the participants in the study (16/20) mostly selected the pictures that represented their own, rather than the speakers's, perspective. If they saw a character localized by the narrator on the left hand side of the screen, they appear to have constructed a mental image in which that character was on the left hand side of the scene or event, as opposed to the right hand side of the scene or event. Only three of the 20 participants were more likely to select the pictures corresponding to the

¹¹Note that 'speaker perspective' is somewhat different in NSL, since signers tend to put themselves into a scene rather than signing as if they are looking at a scene (Senghas and Coppola 2001). For this reason we avoided using the same terms as Senghas and her colleagues for describing perspective.

¹²This is also significantly different from consistent listener perspective responses, (0/12), $t_{2\text{-tailed}}(19) = -5.87$, $p < .0001$.

speaker perspective, and of those, only one was at least 80% consistent in her choices, selecting the listener-perspective picture 10 out of a possible 12 times.

To ensure that our stimuli were not the source of any consistent variation in responses, we examined the data for any patterns based on aspects of the narratives or their gestures and found none. Individual items did not differ from each other in a consistent way: The number of sentences in the narrative did not affect responses, neither did the number of predicates contained in those sentences, nor did the number or nature of the gestures present in the narrations.

In sum, although there is substantial variation, with no participant being completely consistent in his or her responses, there is a significant tendency for participants to adopt their own, rather than the speaker's, perspective.

4. Discussion

We set out to examine the issue of conventionalization in creole languages, pointing out that for a form to emerge as a part of the grammar of the language, it must be interpreted, consistently, and commonly, by most, if not all, users of the language. With respect to space and the particular form-meaning mapping that we examined here, this means that a comprehender must be able to link entities with a spatial location, remember which entities are located where in the abstract signing space, and then recognize those locations across clauses or utterances. Furthermore, the comprehender must interpret the viewpoint consistently for every distinct conversation, that is, the viewpoint taken must be the same between as well as within individual communicative interactions. And lastly, every user of the language must do all of this in the same way as other users. We conducted a study examining consistency within and across people (the latter two requirements we describe), and found no conclusive evidence for either. That is, adults without exposure to a signed language did not interpret the spatial content of referential gestures in a consistent way, individually or collectively (although there was a tendency shared by most participants to interpret things from their own perspective). These results suggest that even under the ideal conditions (i.e., where the interpretation problem is minimized), conventionalization remains a problem to be solved by users of the language.

Our point thus far has been to highlight the difficulty of conventionalization in emerging languages, in particular, to show that people do not automatically converge on the same meaning, with the primary aim of convincing others that it is an issue worth investigating. Indeed, Mufwene recently pointed to the emergence of shared norms as being one of the major distinctions between second language acquisition and creole emergence (Mufwene 2010). But the question of how a form acquires a shared meaning remains. The evidence from NSL suggests that child learners might be important, even crucial, to this process (Senghas 2003). We suggest that this is due to the confluence of three factors. First, children expect linguistic forms to have meaning (Clark 1987, 1993). This is part of the intentionality of language: children assume that we say something because we mean something, and thus, language learning is the process of figuring out how to 'mean' appropriately (Tomasello 2003; Hudson 2004). Second, there is emerging evidence that children are strongly predisposed to conventionality in language (e.g., Graham et al. 2006; Buresh and Woodward 2007; see also Sabbagh and Henderson 2007). That is, they expect linguistic forms to have the same meaning for different speakers and in different contexts, and are rather inflexible about it. Finally, children dislike unpredictability in linguistic forms and will impose their own regularity on inconsistent patterns present in their input (Hudson Kam and Newport 2005, 2009). These three things together mean that they will take a probabilistically occurring form and make it more predictable by conditioning it on linguistic factors (most

likely to be semantic), and will assume that everyone must use the form to mean the same thing.

In the case of NSL, these factors interact to produce conventionalization in the following way. We suggest that for signers in the first cohort, production of spatial modulation is rather like co-speech gesture, with aspects of their mental representations not included in their linguistic productions (in this case because the language does not have a way to express them), leaking out through the hands (McNeill 1992; Goldin-Meadow 1997). As such, the production of spatial modulation would not be expected to be systematic either within or across signers, nor would we expect it to be produced all the time, consistent with the NSL data (Senghas and Coppola 2001). When a new group of children learned from this input, however, they literally saw things quite differently. The spatial modulation was occurring in the same representational format as the language, that is, visual-spatial. They had no means of distinguishing intentional and non-intentional movements, and thus assumed that the spatial modulations were linguistic forms and therefore meaningful (Clark 1987, 1993). They further assumed consistency in usage (Hudson Kam and Newport 2005, 2009) and meaning (Sabbagh and Henderson 2007) – that is, that forms have a consistent relationship to the contexts in which they are used, with respect to both grammar and meaning. In this way an accidental aspect of the first cohort's signing was reinterpreted as part of the language.

The scenario we have just outlined, while explaining how spatial modulation might have become conventionalized in NSL, does not explain how or why the second cohort signers all ended up with the same perspective. There are several possibilities upon which we speculate here. Without more data, however, we cannot distinguish among them in any meaningful way. The first possibility is suggested by the literature on conventionality in children. Research shows that when there are multiple possible interpretations for a novel word, children assume the one they've heard most often (from the most people) is the right one (Corriveau et al. 2009). Thus, it is possible that the most frequent relationship between the world and the spatial modulation experienced by the second cohort was a rotated perspective, that is, that the signers in the first cohort tended to put themselves in the action. Although this does not seem to be the case according to the data presented in Senghas (2003), it could have occurred if, for instance, most of the conversations centered around things the signer themselves actually did or experienced and that this encouraged the signer to sign themselves into the action, resulting in a rotated perspective. Another possibility is that, despite children's tendency to have difficulty taking another person's perspective (Wellman 1990; cf. Onishi and Baillargeon 2005), linguistically, it is the natural thing to do. Tomasello (2003), for instance, argues that mind-reading ability is crucial to learning language. Indeed, according to him it is required if learners are ever to figure out what a form in the language means; they must think about what the other person intends to communicate, which includes thinking about his/her viewpoint. Thus, it may be that when it comes to language the other's viewpoint is the most natural one to assume when trying to assign a meaning to a form.

Although our own results seem to contradict this latter possibility, aspects of our experiment could be to blame for the fact that the listener's perspective that was favoured over the speaker's by our participants. In particular, gestures, although clearly influenced by language and very much integrated with speech and language (McNeill 1992, 2005; Goldin-Meadow 2003a), are not linguistic forms. Importantly, and in contrast to the second cohort NSL learners (who adopted the speaker's perspective), this would be known, at least implicitly, by our participants. Thus, if the push to take a speaker's perspective comes from trying to interpret linguistic forms, it would not automatically apply to the gestures witnessed by our participants, leaving them free to take whatever perspective came more

naturally to them in this context. Another possibility related to the stimuli is that the push to take the speaker's perspective may emerge from the conversational nature of much actual language use. Listeners in a conversation may be more attuned to the fact that a speaker is likely saying things from his/her own perspective, and so more automatically interpret things from the speaker's perspective rather than their own. However, our stimuli were not conversational – there was no interaction between the narrator and the participant – which may have removed the impetus to take the speaker's perspective. Whatever the case with participants in our experiment, one of the interpretations was clearly more natural for the second cohort NSL child learners to assume, and that was the one that emerged.

Importantly, the emergence of shared meanings would occur in much the same way in an emerging spoken language. Children hear a form being used, assume some meaning for the form, assume that all speakers share that same meaning, and proceed to use it accordingly. It is a little more difficult to explain how different children might all end up with exactly the same meaning for an emerging form in the case of a spoken language where there is little or no iconicity to guide learner's hypotheses, as is the case with an emerging signed language. Linguistic constraints may help to at least narrow the range of possibilities (e.g., a form adjacent to a verb likely modifies the verb rather than a noun and so will more likely be interpreted as tense or aspect rather than case), but they will not completely determine potential meanings. We suggest that children's general cognitive limitations and relatively poor theory-of-mind abilities may actually aid in the conventionalization process (Newport 1990; Sabbagh and Henderson 2007; Hudson Kam and Newport 2009); if one meaning is more naturally accessible to children because of their cognitive immaturity, and they fail to appreciate that other people might not share their interpretation, they will rather accidentally share similar interpretations and expect that others will as well.¹³ Because this is an individual-level process, it can only have an impact on a language more broadly if there are a large number of children in the speech community. Thus, conventionalization may not occur immediately. Alternatively, in cases with fewer constraints on possible meanings, individual children may not all end up with the same meaning for a form, only related ones. If so, then some other force would be required to induce speakers to the same interpretation later (see, e.g., Mufwene, 2010). This process would likely take some time to produce convergence within an entire speech community, and so is also consistent with data suggesting that creole languages often take several generations of speakers to normalize (Arends 1993; Mufwene 2010).

In the end, however, this is speculation. All we have truly shown is that conventionalization is not automatic. Indeed, even in situations where it might be assumed to be easy, as in a signed language, it is not. Rather, it is a problem that users of the language must solve. We have offered a possible explanation – that children are agents of conventionalization when new languages emerge – and suggested that this is due to limitations in their abilities, rather than any special endowment they may have for learning language. This is clearly something that begs for more research, and we hope that our efforts stimulate others' interest in this question.

¹³Readers may remark here that children learning established languages do this all the time – hone in on a meaning for a spoken form lacking in iconicity – and thus assume that figuring out the meaning of a grammatical morpheme in an emerging spoken language is not really as much of a problem as we suggest. However, in established languages the meaning of the morphemes are fixed and shared already, not inconsistent and variable. It is rather like the difference between trying to hit a stable versus a moving target. Moreover, even in established languages children sometimes do not initially appreciate the full meaning of a grammatical morpheme (e.g., Bloom et al. 1980), suggesting a) that some meanings may indeed be more accessible or more natural and b) learning the meaning even in an established language can be challenging.

Acknowledgments

We wish to thank Marie Coppola for many discussions about NSL and related themes that inspired this project, and to Darlene LaCharité for helpful comments and critiques on earlier drafts. Thanks also to two anonymous reviewers, whose input led to numerous improvements in this paper. Any remaining errors are our own. This research was supported by NIH grant HD 048572 to C. Hudson Kam.

References

- Arends, Jacques. Towards a gradualist model of creolization. In: Byrne, Frances; Holm, John, editors. *Atlantic meets Pacific: A global view of pidginization and creolization*. Amsterdam: John Benjamins Publishing Company; 1993. p. 371-380.
- Art Explosion. Calabasas CA: Nova Development; 2004. Computer Software
- Bloom, Lois; Lifter, Karin; Hafitz, Jeremie. Semantics of verbs and the development of verb inflection in child language. *Language*. 1980; 56:386-412.
- Brown, Penelope; Levinson, Stephen C. Linguistic and nonlinguistic coding of spatial arrays: Explorations in Mayan cognition. Nijmegen: Cognitive Anthropology Research Group, Max Planck Institute for Psycholinguistics; 1993. Working Paper No 24
- Buresh, Jennifer Sootsman; Woodward, Amanda L. Infants track action goals within and across agents. *Cognition*. 2007; 104:287-314. [PubMed: 16930577]
- Clark, Eve V. The principle of contrast: A constraint on language acquisition. In: MacWhinney, Brian, editor. *Mechanisms of language acquisition*. Hillsdale, NJ: Lawrence Erlbaum; 1987. p. 1-33.
- Clark, Eve v. *The lexicon in acquisition*. Cambridge: Cambridge University Press; 1993.
- Coppola, Marie E. Doctoral dissertation. The University of Rochester; 2002. The emergence of grammatical categories in home sign: Evidence from family-based gesture systems in Nicaragua.
- Coppola, Marie; So, Wing Chee. Abstract and object-anchored deixis: Pointing and spatial layout in adult homesign systems in Nicaragua. In: Brugos, Alejna; Clark-Cotton, Manuella R.; Ha, Seungwan, editors. *Proceedings of the Boston University Conference on Language Development* 29. Vol. 1. Somerville, MA: Cascadilla Press; 2005. p. 144-155.
- Corriveau, Kathleen H.; Fusaro, Maria; Harris, Paul L. Going with the flow: Preschoolers prefer nondissenters as informants. *Psychological Science*. 2009; 20:372-377. [PubMed: 19207691]
- DeGraff, Michel AF. Morphology in creole genesis: A prolegomenon. In: Kenstowicz, Michael, editor. *Ken Hale: A Life in Language*. Cambridge, MA: The MIT Press; 2001. p. 53-121.
- Goldin-Meadow, Susan. When gesture and words speak differently. *Current Directions in Psychological Science*. 1997; 6:138-143.
- Goldin-Meadow, Susan. *Hearing gesture: How our hands help us think*. Cambridge, MA: Belknap, Harvard University Press; 2003a.
- Goldin-Meadow, Susan. The resilience of language: What gesture creation in deaf children can tell us about how all children learn language. New York: Psychology Press; 2003b.
- Goldin-Meadow, Susan. What language creation in the manual modality tells us about the foundations of language. *The Linguistic Review*. 2005; 22:199-225.
- Goodrich, Whitney. Doctoral dissertation. University of California; Berkeley: 2009. Gesture as input in language acquisition: Learning 'who she is' from 'where she is.'
- Goodrich, Whitney; Hudson Kam, Carla L. Pointing to 'her': The effect of co-speech gesture on pronoun comprehension; Paper presented at the 20th Annual CUNY conference on human sentence processing; La Jolla, CA. 2007.
- Goodrich Smith, Whitney; Hudson Kam, Carla L. Pointing to 'her': The effect of co-speech gesture on pronoun resolution. Ms. University of Maryland and University of British Columbia; 2010.
- Graham, Susan A.; Stock, Hayli; Henderson, Annette ME. Nineteen-month-olds' understanding of the conventionality of object labels versus desires. *Infancy*. 2006; 9:341-350.
- Hockett, Charles F. The origin of speech. *Scientific American*. 1960; 203:88-96. [PubMed: 13683472]
- Hudson, Carla L. Review of Michael Tomasello. *Constructing a language: A usage-based theory of language acquisition*. *PsycCRITIQUES: Contemporary Psychology--APA Review of Books*. 2004; 49(Suppl. 14) Article 66.

- Hudson Kam, Carla L.; Newport, Elissa L. Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language Learning and Development*. 2005; 1:151–195.
- Hudson Kam, Carla L.; Newport, Elissa L. Getting it right by getting it wrong: When learners change languages. *Cognitive Psychology*. 2009; 59:30–66. [PubMed: 19324332]
- Kegl, Judy; McWhorter, John. Perspectives on an Emerging Language. In: Clark, Eve V., editor. *Proceedings of the Twenty-eighth Annual Stanford Child Language Research Forum*. Palo Alto, CA: CSLI; 1997. p. 15-36.
- LeFebvre, Claire. The emergence of productive morphology in creole languages: The case of Haitian Creole. In: Booij, Geert; Van Marley, Jap; Plag, Ingo, editors. *Yearbook of morphology 2002*. Part I. Dordrecht: Kluwer Academic Publishers; 2003. p. 35-80.
- Li, Peggy; Gleitman, Lila. Turning the tables: Language and spatial reasoning. *Cognition*. 2002; 83:265–294. [PubMed: 11934404]
- McNeill, David. *Hand and mind: What gestures reveal about thought*. Chicago, IL: University of Chicago Press; 1992.
- McNeill, David. *Gesture and thought*. Chicago, IL: University of Chicago Press; 2005.
- Mufwene, Salikoko S. SLA and the emergence of creoles. *Studies in Second Language Acquisition*. 2010; 32:359–400.
- Newport, Elissa L. Maturation constraints on language learning. *Cognitive Science*. 1990; 14:11–28.
- Newport, Elissa L.; Supalla, Ted. Sign language research at the millennium. In: Emmory, Karen; Lane, Harlan L., editors. *The signs of language revisited: An anthology to honor Ursula Bellugi and Edward Klima*. Mahwah, NJ: Lawrence Erlbaum; 2000. p. 103-114.
- Onishi, Kristine H.; Baillargeon, Renee. Do 15-month-old infants understand false beliefs? *Science*. 2005; 308:255–258. [PubMed: 15821091]
- Padden, Carrol. Doctoral dissertation. University of California; San Diego, CA: 1983. *Interaction of morphology and syntax in American Sign Language*.
- Park-Doob, Mischa A. Doctoral dissertation. University of California; Berkeley: 2010. *Gesturing through time: Holds and intermodal timing in the stream of speech*.
- Plag, Ingo, editor. *The morphology of creole languages*. Special section of *Yearbook of morphology 2002*. Dordrecht: Kluwer Academic Publishers; 2003.
- Plag, Ingo. Morphology in pidins and creoles. In: Brown, Keith, editor. *The encyclopedia of language and linguistics*. 2. Amsterdam: Elsevier; 2006. p. 304-308.
- Pyers, Jennie E.; Senghas, Ann. Reported action in Nicaraguan and American Sign Languages: Emerging versus established systems. In: Perniss, Pamela; Pfau, Roland; Steinbach, Markus, editors. *Visible variation: Comparative studies on sign language structure*. New York: Mouton de Gruyter; 2007. p. 279-302.
- Pyers, Jennie E.; Shusterman, Anna; Senghas, Ann; Spelke, Elizabeth E.; Emmory, Karen. Evidence from an emerging sign language reveals that language supports spatial cognition. *Proceedings of the National Academy of Science*. 2010; 107:12116–12120.
- Sabbagh, Mark A.; Henderson, Annette ME. How an appreciation of conventionality shapes early word learning. *New Directions for Child and Adolescent Development*. 2007; 115:25–37. [PubMed: 17441545]
- Senghas, Ann. Doctoral dissertation. Massachusetts Institute of Technology; 1995. *Children's Contribution to the Birth of Nicaraguan Sign Language*.
- Senghas, Ann. Intergenerational influence and ontogenetic development in the emergence of spatial grammar in Nicaraguan Sign Language. *Cognitive Development*. 2003; 18:511–531.
- Senghas, Ann; Coppola, Marie. Children creating language: How Nicaraguan Sign Language acquired a spatial grammar. *Psychological Science*. 2001; 12:323–328. [PubMed: 11476100]
- Senghas, Richard J.; Monaghan, Leila. Signs of their times: Deaf communities and the culture of language. *Annual Review of Anthropology*. 2002; 31:69–97.
- Senghas, Richard J.; Senghas, Ann; Pyers, Jennie E. The emergence of Nicaraguan Sign Language: Questions of development, acquisition, and evolution. In: Parker, Sue Taylor; Langer, Jonas;

Milbrath, Constance, editors. *Biology and knowledge revisited: From neurogenesis to psychogenesis*. Mahwah, NJ: Earlbaum; 2005. p. 287-306.

So, Wing Chee; Coppola, Marie; Licciardello, Vincent; Goldin-Meadow, Susan. The seeds of spatial grammar in the manual modality. *Cognitive Science*. 2005; 29:1029–1043. [PubMed: 21702801]

So, Wing Chee; Kita, Sotaro; Goldin-Meadow, Susan. Using the hands to identify who does what to whom: Gesture and speech go hand-in-hand. *Cognitive Science*. 2009; 33:115–125. [PubMed: 20126430]

Supalla, Ted. Doctoral dissertation. University of California; San Diego, CA: 1982. Structure and acquisition of verbs of motion and location in American Sign Language.

Tomasello, Michael. *Constructing a language: A usage based theory of acquisition*. Cambridge, MA: Harvard University Press; 2003.

Wellman, Henry M. *The child's theory of mind*. Cambridge, MA: MIT Press; 1990.

Figure 1.
Example localizing gestures.

Figure 2.
Example test pictures (Art Explosion 2004).

Figure 3.
Number of participants (max = 20) selecting the speaker perspective pictures 0-12 times (max = 12).