

MODALLY HYBRID GRAMMAR?  
CELESTIAL POINTING FOR TIME-OF-DAY REFERENCE IN NHEENGATÚ

SIMEON FLOYD

*Max Planck Institute for Psycholinguistics*

From the study of sign languages we know that the visual modality robustly supports the encoding of conventionalized linguistic elements, yet while the same possibility exists for the visual bodily behavior of speakers of spoken languages, such practices are often referred to as ‘gestural’ and are not usually described in linguistic terms. This article describes a practice of speakers of the Brazilian indigenous language Nheengatú of pointing to positions along the east-west axis of the sun’s arc for time-of-day reference, and illustrates how it satisfies any of the common criteria for linguistic elements, as a system of standardized and productive form-meaning pairings whose contributions to propositional meaning remain stable across contexts. First, examples from a video corpus of natural speech demonstrate these conventionalized properties of Nheengatú time reference across multiple speakers. Second, a series of video-based elicitation stimuli test several dimensions of its conventionalization for nine participants. The results illustrate why modality is not an a priori reason that linguistic properties cannot develop in the visual practices that accompany spoken language. The conclusion discusses different possible morphosyntactic and pragmatic analyses for such conventionalized visual elements and asks whether they might be more crosslinguistically common than we presently know.\*

*Keywords:* multimodality, time reference, composite utterances, celestial gesture, pointing, Nheengatú, Brazil

**1. INTRODUCTION.**

**1.1. MULTIMODALITY AND LINGUISTIC SYSTEMS.** In a seminal article, McNeill (1985) argued against characterizing ‘gesture’, or the meaningful visible movement of the hands and body, as ‘nonverbal’, demonstrating many ways in which gesture and speech

\* This article is dedicated to Dona Marcilia Brazão, who taught me so much about language and culture on the Rio Negro, and who kept teaching me new things until the day she passed away, at the end of my last field trip. Additional thanks go to Marcilia’s family members, especially her daughter Lourdes Brazão for consulting on transcription and elicitation materials, and to the entire household of Humberto Silvestre, who hosted me in São Gabriel da Cacheira, and Raul da Silva, Luis Brazão, Silvana Brazão, Bernadete Brazão and Magão and their family, and Elizete Cruz Henrique. Thanks also to the entire community of Castanheiro and all the participants in elicitation sessions in SGC. In Manaus, thanks to Ana Carla Bruno and Aldevan Brazão and family for much logistic and moral support, and also to Denny Moore and INPA/MPEG for the use of vital equipment on my first field trip. Support for fieldwork was supplied by a National Science Foundation Graduate Research Fellowship, the University of Texas, and the European Research Council project ‘Human Sociality and Systems of Language Use’, directed by Nick Enfield and hosted by MPI Nijmegen. Thanks also to Nick for tirelessly commenting on evolving versions of the article, and to many others for very helpful comments on earlier drafts, including Pattie Epps, Aline da Cruz, Connie de Vos, Mark Dingemans, Kensy Cooperrider, and Olivier Le Guen, and thanks for ideas from discussions with Susan Goldin-Meadow, Lev Michael, Eve Clark, and Herb Clark, from Judith Holler and other attendees at a Nijmegen Gesture Center Presentation, from participants in a data session at UCSD organized by Eric Hoenes with John Haviland and Carol Padden, from attendees at presentations at the 2008 Arizona Linguistic Anthropology Symposium, to a session of the 2010 International Society for Gesture Studies conference in Frankfurt, Oder, organized by Kensy Cooperrider, with the participation of Terra Edwards, Isabel Galhano (among others), and to a 2013 Radboud University Nijmegen workshop organized by Josh Birchall, and from conversations in SGC with Sarah Shulist. Thanks to Jeremy Hammond, Gertie Hoymann, and Sabine Reiter for showing me unpublished data on time telling with celestial gestures in their languages of expertise, and thanks for stimulation from many more great colleagues at UT, MPI, and elsewhere, who are too numerous to mention here. Thanks to Francisco Torreira for help with the statistics, and to Rosa Elena Donoso for designing Figure 3. The insightful comments from editor Greg Carlson, associate editor Jürgen Bohnemeyer, and two anonymous referees were extremely helpful during revisions. Any remaining problems or errors are the responsibility of the author.

pattern together in discourse. He proposed that to ‘consider *linguistic* what we can write down, and *nonlinguistic*, everything else’ (1985:350) is an artificial separation of integrated parts of a whole (a position similar to that taken by Kendon (2004, 2008), among many others). Research on multimodality has since continued to feel tension between the need to generate specialized terms and concepts for the visual modality and the basic fact of language usage that meaning is made through combinations of ‘diverse semiotic resources’ (Goodwin 2000), including the articulations of the hands and body, and that interactive moves do not just consist of spoken words but also of multimodal ‘composite utterances’ (Enfield 2009, 2013, drawing on Kendon’s ‘visible action as utterance’ (2004); similar concepts include ‘composite signals’ in Clark 1996 and ‘integrated messages’ in Bavelas & Chovil 2000).

While applying one set of analytical concepts to the visual modality and a distinct set to the auditory modality has led to advances in gesture studies, it has also reinforced the tendency to look just to the auditory channel to study the linguistic system proper, and to look to the visual channel only for more holistic and depictive forms of expression. A nice contrast to this perspective can be found in sign language linguistics, which began its modern era with the realization that language in the visual modality, while perhaps not something that ‘we can write down’ easily with an alphabet designed for spoken language, is robustly analyzable in morphosyntactic terms (Stokoe 1960, Stokoe et al. 1965, Klima & Bellugi 1979, Liddell 2002, 2003). Aside from the exception of ‘auxiliary sign systems’, discussed in more detail below, linguists have not generally expected to see linguistic properties in the communicative visual practices of users of spoken languages. Goldin-Meadow and colleagues (1996) go as far as to predict that populations will only develop such conventionalized properties when the spoken channel is not available, as in deaf communities. It is worth asking, however, whether multimodal practices in spoken languages absolutely NEVER display any of the properties of linguistic systems, or if in some cases these properties may have just gone unnoticed.

This article offers a case study of a phenomenon that speaks directly to this problem: a visual bodily practice used by speakers of the Brazilian indigenous language Nheengatú for time reference. This practice demonstrates the main properties expected from an element of a linguistic system, both in terms of conventionalization, defined by stable form-meaning pairings, and of productivity, defined by systematic oppositions of forms in association with changes in utterance meaning. Its basic elements, described in greater detail in subsequent sections, are as follows:

- When Nheengatú speakers reference a specific time of day, they use a construction that includes simultaneous visual and auditory articulation.
- The auditory element is a verb phrase, minimally including a predicate and optionally including spoken adverbial modifiers.
- The visual element is, minimally, a point indicating the position of the sun at reference time, sometimes indicating two solar positions and the path between them.
- Elements from both modalities combine compositionally according to specific restrictions, the visual element productively adding a conventional, context-independent adverbial meaning.
- When those restrictions are violated, speakers show sensitivity to departures from standards of form (what might be considered a kind of ‘grammaticality’).

My analysis of this practice is guided by Okrent’s (2002) position that auditory and visual elements of utterances should not be distinguished a priori in terms of which are linguistic elements, but should instead be evaluated with ‘modality-free’ criteria based

on semiotic properties. It is important to note from the outset that the way the word ‘gesture’ is used in the literature unfortunately often conflates modality and semiotic properties, so that everything that is ‘gestural’ (i.e. visual) in terms of articulation is also implied to be ‘gestural’ (i.e. ‘nonlinguistic’) in terms of its status as part of a linguistic system. In order to keep these meanings apart, I avoid the word ‘gesture’ when possible, instead mainly using the more specific terms ‘visual’ and ‘auditory’ in reference to modality, and ‘conventionalized’ (or ‘grammaticalized’) and ‘nonconventionalized’ in reference to linguistic systems. Taking care to keep the modality and degree of morphosyntactic conventionalization as two separate questions (Enfield 2009:12–19), we can see how writing systems not only leave out the visual modality altogether, but they also only represent certain elements of spoken language, omitting the ‘unwritable’ elements of speech as well.<sup>1</sup> As illustrated in Table 1, phenomena like Nheengatú time reference, as well as the signs of sign languages, are examples of (i) conventionalized linguistic elements that are (ii) articulated in the visual modality (the shaded cell). Linguistic conventionalization is determined by the extent to which speakers as a group display standards of form, uniform meaning, and productivity, regardless of the modality of expression.

	AUDITORY	VISUAL
MORE CONVENTIONALIZED	spoken elements with linguistic properties (‘writable’ phonemes, morphemes, words, phrases, clauses)	visual elements with linguistic properties (signs, ‘emblems’, conventionalized ‘gesture’ such as Nheengatú time reference)
LESS CONVENTIONALIZED	‘gestural’ elements of speech (‘unwritable’ aspects of prosody, speech rate, volume, voice quality, etc.)	(speech-accompanying) nonconventionalized gesture.

TABLE 1. Modality and conventionalization as two distinct dimensions.

This article makes the case that Nheengatú time reference is a conventionalized system expressed in the visual modality whose properties are more comparable to those of sign languages than to those usually attributed to speech-accompanying gesture. A related proposal, more difficult to decisively determine, is that this system is actually a subsystem of the spoken language, and that the only substantial differences between visual and spoken time references stem from their modality of expression, not from different statuses relative to Nheengatú grammar. As is illustrated in the following sections, both the natural speech data and the results of controlled tests confirm the first proposal, that Nheengatú time reference is, in all important respects, basically a kind of specialized sign language, if one drastically limited in function compared to full-blown sign languages. In many ways this is unsurprising, since sign language research has for decades been documenting the broad human capacity for linguistic expression in the visual modality. As for the second proposal, that Nheengatú visual time reference is part of a larger linguistic system of predicate modifiers also including spoken temporal adverbials, this presents a greater challenge to traditional understandings of the division between grammar and usage in which the contribution of visual bodily behavior to meaning is usually thought to be at the pragmatic level, not at the morphosyntactic level.

<sup>1</sup> It is important to note that the ‘unwritable’ elements of the voice, like intonation, volume, speech rate, voice quality, and so on, have also been excluded from much linguistic analysis, and these elements share some semiotic properties with visual/bodily articulation (see Wilkins 2006:131–32); Clark refers to these as ‘vocal gestures’ (1996:182). Okrent (2002) refers to ‘gestural’ elements in both the visual and auditory modalities.

Given the many different ways that linguists think about grammatical properties, it may not be possible to determine where Nheengatú time reference is appropriately situated in a dichotomy of ‘inside’ or ‘outside’ Nheengatú grammar. Rather than attempting to fully resolve this issue, this article endeavors to lay out some compelling facts about Nheengatú speakers’ visual bodily communicative practices, and then asks what these facts might mean for how we analyze the relationship between grammar and modality. One provocative way of thinking about Nheengatú time reference is that the spoken and visual elements combine in unified constructions that might be thought of as ‘modally hybrid’, in the sense that they draw from two heterogeneous sources to form a whole. This perspective aligns with an emerging body of work that has found ways of applying linguistic analysis to multimodal communication, including that of Slama-Cazacu (‘mixed syntax’; 1976), Fricke (‘multimodal grammar’; 2012), and Jouitteau (‘multi-channel syntax’; 2004a,b, 2007).<sup>2</sup> But research in this area is still developing, so it may be premature at present to evaluate the importance of such findings for our understanding of language. Discovery of more phenomena like that described here for Nheengatú may help to settle some of these issues in the future.

This article proceeds as follows. First it gives a descriptive account of how Nheengatú speakers add time-of-day information to verb phrases using the visual modality, based on examples from a natural speech corpus. Then it subjects Nheengatú time reference to a series of controlled elicitation tests in order to determine whether it shows the properties expected from elements of a linguistic system. One interesting outcome of the development of these tests was the creation of new types of video-based elicitation methods designed for studying the linguistic properties of composite utterances, providing a potential methodological model for paradigms that might be applied to other languages.

**1.2. POINTING AT TIME OF DAY IN NHEENGATÚ.** The basic phenomenon of interest is illustrated in example 1.<sup>3</sup> The speaker is telling a traditional story in which some children are transformed into birds during the night, and then later they are observed by men who arrived the next day. In the example, the verb ‘arrive’ is modified by the adverbial phrase ‘in the morning when the sun is here’. In alignment with the word ‘here’, the speaker points (and aligns her gaze) west at about 65° of altitude to the place where the sun would be at around 11 AM.

<sup>2</sup> Jouitteau (2004a:432–33) emphasizes the wider relevance of language-particular practices that combine auditory and visual morphosyntax:

All spoken languages are potentially multichannel. The arbitrary dimension of the linguistic sign extends to the channel realising particular phonological embodiment. We should then wonder why it isn’t the case that more oral languages use multichannel signs. My guess is that we are not sure they do not. Rarely are the methodologies adjusted to the multichannel dimension of the message. We are in danger of trying to account for truncated data if only the acoustic string is taken into account.

<sup>3</sup> For this and all later examples, the following abbreviations are used: 1/2/3/PL/SG: person/number, AFF: affirmation, ALL: allative, COM: comitative, DIM: diminutive, IPFV: imperfective, IRR: irrealis, LOC: locative, PFV: perfective, PL: plural, Q: question marker, REP: reportive, SUB: subordinator.

Note that color versions of all photo images can be accessed in the HTML version of this article: <http://muse.jhu.edu/journals/language/v092/92.1.floyd.html>.



- (1) Kuémete =paá, [iké kurasí] =paá, u-sika pigã-ta,  
 morning =REP here sun =REP 3SG-arrive man-PL  
 ‘In the morning, they say, [at 11 AM], they say, the men arrived,’  
 ta-wiké =paá fu fu fu fu u-sú u-wapika travesão-pé, aetá.  
 3PL-enter =REP fu fu fu fu 3SG-go 3SG-sit beam-LOC 3PL  
 ‘they entered, they say, and foo foo foo foo they went and sat on the  
 beam.’

During my first experiences transcribing video data with a speaker of Nheengatú, I noticed that my consultant, Luis Brazão, sometimes provided me with more precise temporal information in the Portuguese translation than I could observe in the transcribed Nheengatú speech. For example, for 1 above he said ‘eleven o’clock’ in Portuguese. In my experience consultants sometimes embellish their translations, so I double-checked by asking him how he knew the precise hour if the speaker had not actually mentioned it. He pointed out that the speaker had indeed ‘said’ the time by pointing to a position in the sky. I had made the mistake of fixating on the auditory channel, and I had to learn to ‘listen’ like my consultant did, with the expectation of finding meaningful linguistic elements in the whole multimodal utterance. Now realizing that this information was being encoded in the visual modality, I was able to find many instances of it in the corpus I was recording, always articulated along the east-west axis of the sun’s arc. I assembled a data set of forty-five predicates combined with time-of-day reference produced by five different speakers, which constitutes the main data discussed in §2.

**1.3. CRITERIA FOR LINGUISTIC ELEMENTS.** In order to make a case for Nheengatú time reference as a visual linguistic element similar to the signs of a sign language, it is possible to draw on a number of ‘modality-free’ criteria for distinguishing conventionalized elements of a linguistic system from other elements of multimodal composite utterances. For example, Sweetser (2009) identifies relative degrees of conventionality and analyticity/productivity (in addition to iconicity) as major dimensions of comparison. McNeill (1992:18–23; summarized in Sandler 2003, 2009) offers similar criteria dealing with questions of conventionality and productivity: (i) a ‘gesture’ conveys meaning ‘globally’, while linguistic constructions compose meaning by combining productive subunits; (ii) individual ‘gestures’ do not combine with each other, while elements of linguistic systems combine productively; (iii) conventional form-meaning relationships are more stable for linguistic elements and more context-dependent for ‘gestures’; and (iv) elements of linguistic systems have standards of form, while ‘gestures’ are more idiosyncratic. Goldin-Meadow provides even more detailed criteria for modality-free ‘resilient properties of language’ (2005:2271) that are too extensive to fully address here, but some of the significant criteria include relative degrees of stability, paradigmaticity, categoriality, arbitrariness, and grammatical functionality. The main things we need to know about Nheengatú time-of-day references according to these criteria are whether they productively contribute a stable meaning, whether all speakers comprehend and produce them uniformly, and whether speakers orient to standards of form. The rest of this article applies the criteria above to the Nheengatú data.

**1.4. POINTING AND LINGUISTIC SYSTEMS.** Nheengatú time reference is based on the indexical principles of pointing, harnessing them in the service of a particular construction for time reference. It belongs to a larger family of pointing practices that reflect a basic and prelinguistic human practice of communicatively extending or directing a body part toward a vector in space (Kita 2003, Tomasello 2008, Kendon 2010). One of the major functions of pointing is establishing joint attention, but this is by no means the only way it is used (Enfield and colleagues (2007) discuss distinct pragmatic functions, for example). The diverse functions of pointing arise in part because they become conventionalized in different ways among speakers of different languages. One context in which pointing can be considered obligatory is with many deictic expressions (Kendon 2010), and in addition, research on pointing has shown much cultural variation and meaningful conventionalization of which specific hand shape or body part is used (e.g. Sherzer 1972, Enfield 2001, Wilkins 2003, Cooperrider & Núñez 2012). In such cases it is worth asking to what extent conventionalized pointing practices might be considered part of the spoken linguistic constructions with which they occur. There appear to be many cases beyond Nheengatú in which it could be useful to think in terms of ‘the grammaticalization of pointing’, in the terms Le Guen (2011:296–97) applies in his discussion of conventionalized pointing practices among speakers of Yucatec Maya.

One of the most polarizing debates in sign language linguistics has to do with the linguistic status of pointing for person reference, since it can be difficult to distinguish a more narrow function of pronominal agreement within pointing’s broader function of reference to any point in the surroundings (Liddell 2000, 2003, Lillo-Martin 2002, Meier 2002, Sandler & Lillo-Martin 2005, Johnston 2013). Sign languages may differ in this respect, but there are attested cases of sign languages in which pointing has taken on such a high functional load in the morphosyntax that it seems incorrect to attempt to exclude it from the linguistic system. De Vos (2012, 2015) shows how pointing forms part of a number of subsystems in the Balinese village sign language of Kata Kolok, including pronouns, locatives, color terms, and—most interestingly for the present discussion—a system of celestial pointing for time reference like that of Nheengatú. De Vos argues that many of these pointing practices display the properties of linguistic signs, with varying degrees of ‘morphemization’ of conventionalized forms and ‘syntactic integration’ into specific grammatical slots. In her discussion of Kata Kolok (2015), de Vos suggests that if similar properties could be identified for pointing practices accompanying spoken language, an analogous analysis could apply. The Nheengatú data provide a nice case for comparison, as both languages tell time in remarkably similar ways. In the context of a sign language, such conventionalized and linguistically productive pointing practices are usually treated as elements of the linguistic system. Given the similarity of the practices, it seems possible to approach Nheengatú in a similar way to the comparable practices in Kata Kolok or other sign languages.

**1.5. OTHER PHENOMENA IN THE VISUAL MODALITY.** The literature on gesture and sign language typology describes several kinds of bodily visual expression that can be compared and contrasted with Nheengatú time reference, including metaphorical time gestures, conventionalized ‘emblem’ gestures, the ‘code blending’ of bimodal bilinguals, and auxiliary sign languages used by some hearing populations. Speakers of languages around the world point along spatial axes to map the flow of time (Cooperrider & Núñez 2009; the Aymara ‘backwards future’ is one well-known case (Núñez & Sweetser 2006); other languages have no specific timeline, for example, Yucatec Maya (Le Guen & Pool Balam 2012)); Nheengatú time reference is not a metaphor for time but

instead relates specific times of day to specific solar positions.<sup>4</sup> ‘Gestures’ with conventionalized ‘lexical’ form-meaning pairings have been referred to with the terms ‘emblem’ or ‘quotable gesture’ (Efron 1972 [1941], Ekman 1976, McNeill 2000, Kendon 2004, Payrató 2008); unlike Nheengatú time reference, ‘emblems’ are generally thought of as stand-alone signs with no combinatory properties, although researchers may have not yet fully taken into account the wide cultural variation of conventionalized gesture (documented by Sherzer 1991, Kendon 1992, Hanna 1996, Brookes 2004, 2005, 2011, among others), meaning that some ‘emblems’ may turn out to be more complex than expected.<sup>5</sup> Another multimodal practice that is partly comparable to Nheengatú time reference is ‘code blending’, when bimodal bilinguals who use both a spoken and a signed language sometimes mix both on-line during speech production, each modality complementing the other (Emmorey et al. 2005, Emmorey et al. 2008, Lillo-Martin et al. 2010); the difference compared to Nheengatú time reference is that ‘code blending’ combines two full-blown linguistic systems. Finally, it is worth mentioning ‘auxiliary’ or ‘alternative’ sign languages such as the Plains Indian Sign Language used between people with no spoken language in common (Mallery 1880, Clark 1885, Tomkins 1969 [1931], Davis 2010) and Australian Aboriginal sign languages used to circumvent speech taboos (Kendon 1988a). With respect to Aboriginal sign language, Kendon (1987, 1988a,b) notes that when it is used together with speech it mirrors the content and structure of the spoken language, rather than adding nonredundant information in a morphosyntactic sense.

Neither Plains Indian Sign Language nor Australian Aboriginal sign languages are fully comparable to Nheengatú time reference because both are total alternatives to speaking (motivated by factors like lack of a shared language or the need to avoid speaking due to a taboo), and they do not provide additional meaning in the linguistic/semantic sense when used with speech. Nheengatú time reference, in comparison, occurs together with speech, but cannot be substituted by the auditory resources of the language. What all of these phenomena have in common, however, is that they are based on hearing populations’ conventionalization of form-meaning pairings in the visual modality. It is this possibility for sign-language-type conventionalization of visual bodily articulations together with spoken language that makes possible the development of the multimodal constructions described in detail in the following sections.

## 2. BIMODAL TIME REFERENCE IN NATURAL SPEECH DATA.

**2.1. BACKGROUND: THE LANGUAGE AND THE FIELD SITE.** Nheengatú is a Tupí-Guaraní language descended from the coastal language Tupinambá, adopted as the *língua geral* or ‘general language’ of the Portuguese colony in the seventeenth and eighteenth centuries. It has gone through a number of transformations over the centuries and has come to be called ‘Nheéngatú’ or ‘the good language’ in its modern form. No longer spoken on the coast, it expanded far into the Amazonian interior, where it continues to be spoken today in the Rio Negro area of Brazil and adjacent Colombia and Venezuela (Taylor 1985, Rodrigues 1986, 1996, Moore et al. 1994). The language has

<sup>4</sup> Some languages do use the east-west course of the sun as a general temporal metaphor, as in the Australian Aboriginal community of Pormpuraaw, as described by Boroditsky and Gaby (2010); while I have not systematically studied metaphoric timelines in Nheengatú, I have observed some instances in which past time references cooccur with gestures aimed behind the speaker, with the future ahead.

<sup>5</sup> In a South African case described by Brookes (2004), different articulations of a gesture for ‘clever’ have a stative copular meaning (‘he’s clever’) or an imperative meaning (‘be clever/watch out!’), a system of what might be analyzed as limited inflection or paradigmaticity.

recently been described in depth by da Cruz (2011), whose work informs my analysis and glosses. In the Nheengatú-speaking community where I collected the corpus, the villagers came from many different Tucanoan and Arawakan ethnic lineages (*etnias*) that share elements of local culture and intermarry exogamously between groups. Several generations of language shift to Nheengatú have led to the curious situation of ethnic members of these two language families speaking a Tupí-Guaraní language that is unrelated to their heritage languages (see Floyd 2007 for a more detailed ethnographic account).

The Rio Negro region is famous for extreme multilingualism and practices of linguistic exogamy (Silva 1962, Sorensen 1967, Jackson 1983, Stenzel 2005), and the Nheengatú speakers I worked with can be considered to be part of this general cultural complex. This raises the question of whether this type of temporal reference system is just used by speakers of Nheengatú or whether it is a more general feature of this area of intense language contact. Speakers of other languages in the region can be observed using similar systems, and one of my consultants—a Nheengatú-Tucano bilingual—also was able to make visual time references along with her Tucano speech, so it appears it may be another shared linguistic trait of a region well known for areality effects (Aikhenvald 2002, 2003, Epps 2005, 2009). More generally, the consistent year-long twelve-hour days of the equatorial environment of the Amazon may provide support for the spread and maintenance of time-telling strategies that rely on the sun's arc, since it is a common-ground experience that is easy to exploit, as compared to the variable arcs seen in more northern or southern latitudes where the development of a comparable system would be less likely.

Indeed, there appears to be a strong relationship between these kinds of systems and tropical populations, since while searching the literature and viewing colleagues' video data I have identified many attestations of similar systems around the world, nearly all of which are in the tropics. These include the Yéli Dnye language of Papua New Guinea (Levinson & Majid 2013:4–5); Guugu Yimithirr of Australia; Tzotzil Maya of Mexico (Haviland 2000:17); Yucatec Maya (Le Guen & Pool Balam 2012:3–8), Bororo (Fabian 1992:87–90), and Awetí (Reiter 2012) of Brazil; Kuuk Thaayorre of Australia (Boroditsky & Gaby 2010); White Sands of Vanuatu (Jeremy Hammond, p.c. 2010); and †Akoë Hai||om of Namibia (Gertie Hoymann, p.c. 2012).<sup>6</sup> Additionally, comparable systems have been found in sign languages, such as Kata Kolok, mentioned previously (de Vos 2012, 2015), and Plains Indians Sign language, about which Tomkins wrote in 1931: 'For time of day, make sign for Sun, holding hand toward the point in the heavens where the sun is at the time indicated. To specify a certain length of time during the day, indicate space on sky over which the sun passes' (1969 [1931]:8).

It is unclear in what ways and to what extent visual time reference is conventionalized in all of the languages listed above, but at any rate, different variants of the practice observed with speakers of Nheengatú can be found all over the world.

Nheengatú's complex history makes it difficult to delimit the language's speech community, but the data presented here represent a more-or-less uniform variety of Nheen-

<sup>6</sup> One 1920 source on 'primitive time-reckoning' (Nilsson 1920) compiles references to many more languages and cultures who have similar practices (using the original terms and spellings): the peoples of southern Nigeria, speakers of Swahili, the Wagogo, the Loango, the Masai, 'the Hottentots, who express with certainty and clearness both points and duration of time by referring to the position of the sun' (1920:18), the people of Dahomey, the Caffre, the people of the New Hebrides, the Bontoc Igorot of Luzon, the Kanyans of Sarawak, various people of Java and Sumatra, Tahitians, Torres Strait peoples, the Omaha of North America, and the Karaya of Brazil.

gatú spoken along a stretch of the Middle Rio Negro between the two towns of Santa Isabel do Rio Negro and São Gabriel da Cachoeira, founded as Salesian missions and today the seats of two adjacent administrative divisions. The varieties spoken upriver from São Gabriel, mainly by Arawakan peoples like the Baniwa and Werekena, are distinct from, but largely intelligible with, the downriver varieties.<sup>7</sup> Today Nheengatú is spoken, sometimes along with other languages, in many different villages in different indigenous reservations along the banks and islands of the broad Rio Negro and its many tributaries, as well as in the larger towns, which were historically Nheengatú-speaking centers until Portuguese gained currency in the region (see e.g. Figure 1). These towns have grown in recent years, linked to growth of the Brazilian military installations in the region as well as commercial activities like gold mining, and they have received much migration from the smaller communities in the area.



FIGURE 1. A processing house for the staple food, manioc, in a Nheengatú-speaking village (left). A day's travel upriver is the town of São Gabriel da Cachoeira, where elements of mainstream Brazilian society can be found, such as *lanchonetes* (snack bars) like that shown in the photo (right); this *lanchonete* takes its name from the Nheengatú word for 'sun'.

The natural speech data considered in this article were collected during two field trips to the same Rio Negro community in 2004 and 2007, when I recorded a corpus of traditional stories, personal narratives, conversation, and interviews. Much of these data were transcribed while I was in the community and later while I was working with community members in São Gabriel. While I discovered early on in this work that meaning was sometimes being encoded in the visual modality (as recounted in §1.2), it was difficult to extend traditional elicitation techniques into the domain of multimodality because these are usually based on constructions that can be written down and repeated back. Speakers were able to share some metalinguistic judgments during ethnographic interviews; in an interview a speaker named Raul compared the Portuguese and Nheengatú time-reference systems with the compelling statement 'our clock is the sun' ('o relógio nosso é o sol'). Raul's mother-in-law Marcília,<sup>8</sup> the speaker shown previously in 1, was also present, and she and Raul gave detailed explanations of how time references are produced and comprehended: '(someone) arrives when the sun is here [pointing west], marking it like this' ('usika iké ramé kurasí, yawé umarcari'), to which Raul replied, 'marking it (so that) the other guy understands, like a clock' ('umarcari, o cara está sabendo, tipo relógio mesmo,

<sup>7</sup> Two speakers of upriver varieties—one whose first language was Baniwa—participated in the elicitation exercises presented in §3, but their responses were excluded from the data set because dialect differences were significant enough to complicate their ability to make straightforward judgments (although both recognized very similar practices for time reference in their varieties).

<sup>8</sup> Both interviewees are native Nheengatú speakers, although their relationship is a good illustration of the sociolinguistic complexities of the region: Marcília is ethnically Tucano, and speaks both Nheengatú and Tucano (but little Portuguese), while Raul is ethnically Dessano—and as such was able to exogamously marry Marcília's daughter, an ethnic Piratapuya from her father's lineage—and is bilingual in Portuguese and Nheengatú.

ne?’). However, eliciting judgments was not usually this straightforward, because, unlike with monomodal examples that can be read back verbatim or modified systematically for speaker evaluation, my attempts at doing this bimodally had me scrambling to simultaneously speak the language, point with the proper orientation, and syntactically align the two (Figure 2). Particularly with the ‘starred examples’ that cannot be found in natural speech, my status as a nonnative speaker makes it difficult to control which of the potentially many ill-formed elements might affect speakers’ judgment. For this reason I supplemented the 2004/2007 natural speech data with new data from a 2012 field trip in which I applied a video elicitation stimulus series featuring a native speaker. Using this new paradigm I was able to more successfully elicit speaker judgments where traditional elicitation methods had not been as effective. Researchers working with other phenomena involving conventionalization in the visual modality might also in those cases profitably apply a variant of this paradigm, which is described in §3. Before discussing the controlled tests, however, in §2.2–2.4 I first expand on the description of the basic phenomenon as it appears in the natural speech corpus.



FIGURE 2. Interview with Raul and Marcilia in which I attempted to produce multimodal examples for evaluation.

**2.2. VISUAL TIME REFERENCE IN NHEENGATÚ.** Until the other languages with visual time-telling systems mentioned above are studied in more detail, it will be difficult to know to what extent their practices resemble those of Nheengatú speakers. In terms of the expression of temporality in language, however, Nheengatú has one potentially relevant difference from the other languages in the region: while many of these have large tense paradigms, Nheengatú does not obligatorily mark tense, and has limited or no tense marking, depending on one’s analysis.<sup>9</sup> This means that, as Smith (2008) has pointed out for tenseless languages in general, its explicit time references are primarily achieved through adverbial modifiers rather than through verbal marking (see Bohne-meyer 2009, for example, on how this works in Yucatec Maya). Nheengatú complements spoken adverbs with its visual bodily system for specifying time of day. In the spoken language, there are three main daytime adverbs for ‘morning’, ‘afternoon’, and ‘midday’ (referring not to noon sharp but to a period more or less between 11 AM and 1 PM). Celestial pointing, either alone or together with these adverbs, helps to narrow down the referential granularity to about the level of a twelve-unit day (Figure 3). This same level of precision is not available for nighttime reference; speakers neither produced such forms nor did they accept them in elicitation.<sup>10</sup>

<sup>9</sup> Tonhauser (2011) analyzes Nheengatú’s close relative Guaraní in a similar way. There is one marker in Nheengatú that might be analyzed as an immediate future tense, as da Cruz describes it (2011:331–43); however, this marker is not obligatory for future time reference.

<sup>10</sup> The Balinese sign language *Kata Kolok*, which also uses celestial pointing for time reference, presents an interesting case of such a system extending into nighttime references (de Vos 2012). This is not accepted by Nheengatú speakers, who described ways for reckoning time by the moon or stars, but said they were much less usual or exact.

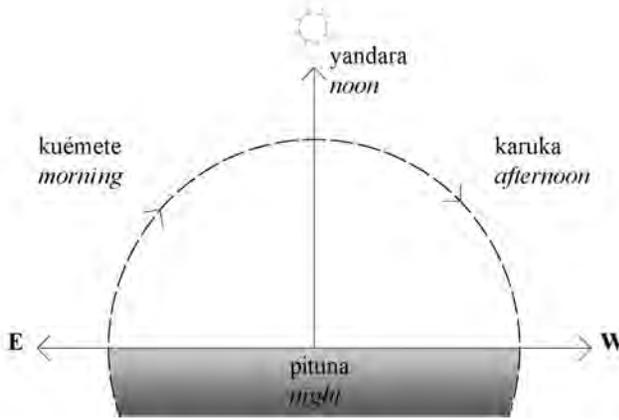


FIGURE 3. The east-west representation of the sun’s arc as observed in the gestures of Nheengatú speakers (extending only slightly into nighttime), with corresponding spoken adverbs at a coarser level of granularity.

Example 2 begins to illustrate the variety of types of cross-modal constructions observed in the corpus, showing a series of different time references in a sequential narration that covers an entire day. Some of the references are for specific moments and consist of stationary points, while others are for periods of time and sweep between two locations. Elements in the spoken line that are simultaneous with visual time references are marked with brackets.



[POINT ‘it got light’] [POINT the ‘sun (is) still here’] [SWEEP> ‘until’][POINT ‘just here’]

- (2) té mamé [yande-kuéma] ya-xari lanterna.  
 until where 1PL-dawn 1PL-leave lantern  
 ‘Until where [it got light at 6 AM], we left the lantern.’  
 [Ike-ré kurasí] ya-sika piasá tía-pe.  
 here-IPFV sun 1PL-arrive palm.fiber grove-LOC  
 ‘(When) [it is still 10 AM] we arrive at the palm fiber grove.’  
 Ya-munuka [até ike-ntu-té], ta-nheé pacote fardo ta-nheé =wasú.  
 1PL-cut until here-only-AFF 3PL-say package 30kg.bag 3PL-say =big  
 ‘We cut [from 12 PM until 6 PM] the package called “fardo”, a big (30 kg) bag.’



[POINT ‘around noon’][SWEEP>] [SWEEP>] [POINT ‘six o’clock’]

Aé aá =rupí ya-yuiri, yawé waá [yandara =siwara], yawé ...  
 3SG there =around 1PL-return like.this SUB noon =since like.that  
 ‘He (the worker, goes) around there and returns, like that [around 12 PM to 6 PM],’

Ya-sika [seis horas], barraca =kití.  
 1PL-arrive six hour thatch.house =ALL  
 ‘like that, we arrive [at 6 PM] at the hut.’

The time references occur together with specific types of spoken material, generally with predicates (like *yande kuéma* ‘it got light’) and predicate modifiers, often with deictic words (*ikeré kurasí* ‘(when) the sun is still here’). While spoken adverbs usually modify through linear contiguity with a predicate, these time references have the possibility of simultaneous expression with a predicate or other spoken material. One way of looking at these time references is as one of several resources that combine to determine the temporal and aspectual semantics of the verb phrase (the ‘verb constellation’; Smith 1997). The two distinct ways of performing time reference—as a stationary point or as a sweep across the sky—contribute two different aspectual meanings: punctuality versus durability. When events are construed as instantaneous, such as killing a monkey in example 3 below, they tend to combine with stationary pointing (here with an open hand shape—possibly related to more general time reference versus a more precise one).



[POINT ‘as it was getting into afternoon’]

- (3) Yawé waá [karuka banda-kití-ã] u-yuká yepé makako.  
 like.that SUB afternoon side-ALL-PFV 3SG-kill one monkey  
 ‘So like that, [as it was getting into afternoon at 3 PM], he killed a monkey.’

In contrast, when events are construed as having inherent duration, such as waiting until noon or dancing all afternoon, as in examples 4 and 5, they tend to combine with sweeping movements indicating a prolonged period of time.



[<SWEEP He ‘waited’] [<SWEEP ‘until’] [POINT ‘noon’]

- (4) U-saarú =paá [até yandara].  
 3SG-wait =REP until noon  
 ‘[He waited from 6 AM], they say, [until 12 PM].’



[<SWEEP ‘they dance’] [<SWEEP] [<SWEEP]

- (5) [Ta-purasí.]  
 3PL-dance  
 ‘[They dance from 10 AM to 6 PM.]’

Combinations of different types of predicates with different time references can derive new ‘situation types’ (Smith 1997), as in example 6 in which a verb that is more often used punctually, ‘to cut’, is used with a durative time reference in order to get a semelfactive reading of multiple punctual events over a period of time: ‘We cut from afternoon until 6 PM’.

(6) Ya-munuka até ike-ntu-té. [SWEEP]

1PL-cut until here-just-AFF

‘We cut [from 12 PM until 6 PM].’ (from example 2; see images above)

Gesture studies have pointed out that regular variation in the form of a gesture can sometimes be associated with regular changes in meaning (Brookes 2001, Haviland 2003, Kendon & Versante 2003, Wilkins 2003), so there is some precedent for finding meaningful ‘inflection’ in the visual elements.

**2.3. POINTING: ORIENTATION AND ARTICULATION.** When speakers of Nheengatú talk about space, most of the time they use linguistic resources that identify locations by way of a fixed axis based on the direction of the Rio Negro, an obvious landmark as one of the biggest rivers in the world, making use of the geocentric directional words *gapira* ‘upriver’ and *tumasá* ‘downriver’. Languages vary with respect to which frame of reference speakers use dominantly (see chapters in Levinson & Wilkins 2006), and speakers may also switch between frames in different contexts or combine them (e.g. Yucatec Maya; Bohnemeyer 2011). In contrast with languages whose speakers tend to rely on the egocentric ‘relative’ left/right frame or the ‘intrinsic’ frame that locates places relative to objects, Nheengatú speakers tend to use an ‘absolute’ system, in the terms of Levinson’s (2003) typology of frames of reference in language, which locates places according to fixed axes like north/south or, in this context, upriver/downriver.<sup>11</sup> When users of geocentric axes like the speakers of Nheengatú (or signers of Kata Kolok; de Vos 2015) adapt long-distance direct pointing to different linguistic functions, correct orientation of pointing becomes a kind of formal restriction on meaningful communication. In the case of time of day, only indicating the sun’s true path in the sky can provide an accurate reference, and speakers go out of their way to do so, as in 7 in which Marcilia twists her body and points backward over her shoulder in order to achieve the proper eastward orientation.



[POINT ‘here sun’]

(7) [Iké kurasí.]  
here sun  
‘[9 AM.]’

<sup>11</sup> Le Guen (2011) shows how each of these frames of reference has different implications for pointing practices; for example, pointing according to the principles of the relative frame of reference means transposing the origo onto the body of the speaker, while pointing according to an absolute frame means that the vector should locate the actual position of the referent in the world. Le Guen calls the latter ‘direct’ pointing; he notes that languages and cultures do not differ with respect to whether they use direct pointing, since it is employed by people everywhere for locating referents in the immediate environment, but instead that there is variation with respect to how far away direct pointing can extend out of the immediate visible surroundings (2011:269).

Accounts of pointing practices from around the world describe a great deal of formal diversity in terms of which hand shape or body part is used to signal the vector (see Sherzer 1972, Haviland 1993, 2000, Enfield 2001, 2005, 2009, Kita 2003, Wilkins 2003, Enfield et al. 2007, Kita 2009, Cooperrider & Nuñez 2012). Nheengatú time reference is usually accomplished with a manual point, frequently accompanied by eye gaze,<sup>12</sup> with the opposition of an index finger versus an open hand potentially linked to the precision of the reference (as noted in §2.2). However, in certain cases it can also be accomplished through gaze alone, as in example 8 in which the speaker's hands were busy because he was weaving a basket while speaking.



[POINT 'this hour']

- (8) *Aqui, amú =ramé ya-mbaú bem cedo, ne?* [Kwaá hora.]  
 here other =when 1PL-eat very early right this hour  
 'Here sometimes we eat really early, right? [At 6 AM.]'

Household objects like tools can also be used for articulation, as in example 9 in which the speaker makes a time reference with the butt of her machete, which she is using to peel manioc.



[POINT 'morning']

- (9) *Wirandé [kuémeté] kuri ya-semu ya-sasá ya-sú São Gabriel =kití.*  
 tomorrow morning IRR 1PL-leave 1PL-pass 1PL-go São Gabriel =ALL  
 'The next day in the [morning at 7 AM] we were going to leave for São Gabriel.'

Pointing practices can be conventionalized in different ways, and in some cases formal elements like hand shape can be standardized (Haviland 2003, Kendon & Versante 2003, Wilkins 2003), and so have the possibility of being ill-formed when they do not meet the standard.<sup>13</sup> Nheengatú time-of-day reference need not necessarily be articulated with the hands, but it must necessarily indicate a vector on the east-west axis. If a time reference were attempted by pointing outside of the east-west axis, or by sweeping backward from west to east, it would no longer be intelligible in terms of time of day and as such would be 'ungrammatical' as a proper time reference. In contrast, common deictic pointing can be directed anywhere in the surroundings and usually refers

<sup>12</sup> Gaze behavior provides some limited evidence that can distinguish pointing in order to establish joint attention on a point in space from other kinds of pointing, since, in the recordings in which it is observable, addressee gaze does not follow the vector of the point but rather stays on the speaker; similar evidence has been applied to pointing practices in signed languages to study how they are comprehended (Pizzutto & Capobianco 2008).

<sup>13</sup> In addition to grammatical ill-formedness, certain kinds of pointing can be culturally prohibited, such as pointing with the left hand (Kita & Essegbey 2001); breaking this norm would be inappropriate, but presumably not unintelligible.

to concrete locations and objects, with the specific relevance of the vector heavily determined by context; it may refer to a place, a person, a trajectory, or any number of other things. Nheengatú time reference has one specific meaning—time of day—and is restricted to a drastically more limited set of vectors whose specific meanings are context-independent.

During the same interview with Raul and Marcilia discussed above in §2.1, Raul made this exact point, showing how east-west pointing gives a time reference while north-south pointing does not: ‘It has to be just from here’ (‘Tem que ser só daqui mesmo’), he said, pointing east, then sweeping east to west: ‘from here to here’ (‘daquí ... para acá’). Then, sweeping north to south, he said, ‘not this direction’ (‘para acá não’). Marcilia added that pointing north to south is possible, but only when used as a deictic directional point and not for time reference. She gave an example of something one could say while pointing north: ‘I’m going, I’m going in that direction’ (‘ixé asú, asú aikú kwaá kití’); ‘then it works’ (‘aape umeé’), she said. Marcilia’s intuitions hold up with the responses to the video stimulus task discussed below in §3. In cases when I intentionally instructed the speaker in the stimulus to point outside of the east-west axis, one common response was to seek an interpretation as a location or direction. If they are asked to limit their interpretation to time of day, however, speakers judge such examples as unsuccessful. While there are some obvious similarities to be found between pointing for time reference and pointing more generally, the stable form-meaning pairing of Nheengatú time reference and its more restricted standards of form contrast with deictic pointing’s freer orientation.

**2.4. AUDITORY/VISUAL COMBINATORY PROPERTIES.** This section completes the description of the Nheengatú time-of-day reference system by making a few last points about its combination with the spoken elements of the composite utterance. It should be noted that bilingual speakers of Portuguese can use the European-style twenty-four-hour system together with celestial pointing, although this occurred only rarely in the corpus. Using both systems together is partly redundant in terms of clock time, since it gives similar temporal information in both modalities at once;<sup>14</sup> in 10 Raul combines a visual time reference with the phrase ‘four in the afternoon’ in Portuguese.



[POINT at four in the afternoon]

- (10) U-sú-ã =riré =paá [quatro horas da tarde], tigre u-uri yawaraté  
 3SG-go-PFV =after =REP four hours of afternoon tiger 3SG-come jaguar  
 =wasú, ne?  
 =big right  
 ‘After he’d gone, they say, [around 4 PM in the afternoon], a tiger came,  
 big jaguar, right?’

Despite the availability of the Portuguese terms, speakers rarely combined these with redundant visual time references, and instead visual time references were articulated si-

<sup>14</sup> Combining Portuguese and Nheengatú resources may be communicatively meaningful in other ways, however; for instance, it appears that open-hand pointing may communicate a less exact time reference (‘right at that time’ versus ‘around that time’), and if this is the case then the open hand shape in 10 may downgrade the precision of the Portuguese time reference ‘four in the afternoon’.

multaneously with more complementary elements. The next part of this section discusses the typology of the different composite utterances combining spoken material and visual time reference observed in the natural speech data, and the different ways they exploit the possibility of simultaneous articulations in two modalities to densely package information (see Enfield 2009:108–9).

There is debate about the implications of small numeral inventories in Amazonian languages for the numerical cognition of their speakers (Gordon 2004, Everett 2005, Frank et al. 2008), but these questions are often motivated by the spoken language alone. For such languages it is also worth considering the ways the visual modality may be used for abstract numerical functions like measuring units of time that in other languages might be accomplished with spoken numerals. Nheengatú speakers can convey precise temporal references simultaneously with spoken elements without competing for a slot in the spoken channel. Looking at the spoken elements alone it might appear that a lack of higher numbers implies the inability to express complex information about temporal duration, without considering the many different ways spoken elements combine with and are complemented by visual elements.

Example 11 shows an extended stretch of discourse describing the three-day process of manioc processing, and includes a variety of time references that exemplify a number of the possible cross-modal construction types. One of these (line 2) is made up of a complex time reference with scope over four predicates in a serial predicate construction. For a spoken language to express such detailed information would require an adjunct phrase like ‘from noon until six’ in the auditory channel.



[POINT ‘morning’] [SWEEP>] [POINT ‘when (the sun is) here’] [SWEEP> 4 verb series]

- (11) Re-karai [kuémete] re-karai, [iké =ramé] re-mbaã,  
 2SG-scrape [morning] 2SG-scrape here =when 2SG-finish  
 ‘You scrape in the [morning at 6 AM], scrape, and [at 12 PM], you finish.’  
 [re-kitika, re-yamí, misturari, re-tipika], pronto.  
 2SG-grate 2SG-squeeze.out mix 2SG-press ready  
 ‘[You grate (manioc), squeeze it, mix it, (then) you press it from 6 AM to  
 6 PM], ready!’



[POINT ‘morning’] [SWEEP> ‘at this time’][POINT ‘there, right there’]

- [Kuémete], yepeã aikwé pã, re-puú pã re-munhá tatá, re-yapí-ã  
 morning firewood exist all 2SG-gather all 2SG-do fire 2SG-throw-PFV  
 re-ikú.  
 2SG-to.be  
 ‘In the [morning from 8 AM to 12 PM], firewood, gather it up, make a fire  
 and throw it in.’

Pu [kwaá hora-upé] forsa, boa-té [mi mi-mirí] re-mbaã,  
 wow this hour-LOC force big-AFF there there-DIM 2SG-finish  
 ‘Wow, [from 12 PM to 6 PM], (work) hard, and [right at 6 PM], you finish.’  
 karuka, pituna-irũ re-mbaã re-ikú uí.  
 afternoon night-COM 2SG-finish 2SG-to.be manioc.flour  
 ‘In the afternoon, around nightfall, you are finishing the manioc flour.’

The example above illustrates how the time references combine with different elements of the spoken line in several different ways. The three general kinds of spoken elements that occur simultaneously with visual time reference are: (i) deictic words that require indexical referents, provided by celestial pointing, (ii) general time adverbs whose precise meaning is narrowed down by finer-grained visual time references, and (iii) predicates where the visual time reference is the sole modifier, simultaneously occurring with a predicate without the support of other adverbial elements (Table 2). Adverbial modifiers generally precede verbs, and when spoken adverbial elements are present, celestial pointing occurs simultaneously with them, occupying the same syntactic position as modifiers in the verb phase. When spoken adverbial elements are not present, this allows for the possibility of expressing the time reference simultaneously with the predicate, an affordance of the visual modality not possible for spoken elements.

	◀ BIMODAL INTERDEPENDENCE	INDEPENDENT TIME REFERENCE ▶	
COMPOSITE UTTERANCE TYPE	(POINT + DEICTIC) + (VERB)	(POINT + ADVERB) + (VERB)	(POINT + VERB)
	<b>deictic construction</b>	<b>complementary adverbs</b>	<b>direct modification</b>
FORM	spoken deictic with visual celestial point  ex. <i>iké ramé</i> (12 PM) ‘here sun’	spoken adverb with visual celestial point  ex. <i>kuémete</i> (6 AM) ‘morning’	spoken predicate with visual celestial point:  ex. <i>rekitika</i> (12 to 6 PM) ‘you scrape’
CROSS-MODAL RELATION	Time reference through indexical word/pointing interdependency; bimodal adverbial phrase adds temporal information to VP.	Time reference through complementary cross-modal elements; bimodal adverbial phrase adds temporal informa- tion to VP.	Visual time reference alone adds temporal information to VP.

TABLE 2. Types of composite utterance observed for Nheengatú visual time reference.

All of the construction types described in Table 2 are in current usage; however, they may represent different stages of grammatical incorporation of the visual elements. Since the deictic constructions resemble deictics more generally, it seems possible that the other composite utterance types started with spoken deictic components and then eventually became less dependent on spoken elements, giving rise to more abstract independent constructions for visual time reference. When occurring with spoken adverbs, celestial pointing is not explicitly linked to a deictic word, but its cooccurrence with a time adverb provides a cue that the intended meaning is temporal. When a visual time reference directly modifies a predicate, it is in its most independent, grammaticalized form, in which it is produced and understood unambiguously as a time reference with no cue from the spoken elements. In such cases, the key to the visual time reference’s appropriateness is that it overlaps with a predicate and that it orients to the east-west axis of the sun’s path (otherwise it cannot be distinguished from other types of pointing; see §§3.2–3.3).

Comparing the different degrees of interdependence between the auditory and visual elements of time reference in the context of the general gestural practices of the speech community suggests a scenario in which practices with less conventionalized (more ‘gestural’) properties developed into practices with more conventionalized, language-like properties. At first, speakers would apply their repertoire of deictic constructions for time reference in specific instances and, finding it a useful affordance of pointing, would begin to habitually use it for further time references until it lost its dependence on context and began to be applied without deictic words. At this point it would have acquired a uniform temporal meaning understood by the members of the speech community (see §3.4) and more rigid standards of form (see §3.3). Sign language researchers have identified similar stages in the development of sign languages, noting that linguistic signs can often be traced back to other kinds of nonlexicalized gestures—including those used in the surrounding hearing communities—that have become increasingly abstract and conventionalized (Kegl et al. 1999, Meier et al. 2002, Goldin-Meadow 2005, Goldin-Meadow et al. 2007, Delaporte & Shaw 2009). Similar processes appear to be at work here, with more general types of pointing becoming conventionalized in new ways for a specific function.

Taking a modality-free approach to these natural speech data appears to reveal a conventionalized subsystem of form-meaning pairings in the visual modality that closely resembles something that might be observed in a sign language. This conclusion still leaves questions about this phenomenon, however. In order to further test this practice to determine the extent and nature of its language-like properties, the following sections present the results of an adaptation of traditional field elicitation methods, here applied to the visual rather than the auditory channel.

**3. A FIELD ELICITATION METHOD FOR COMPOSITE UTTERANCES.** Looking for language-like properties in a speaking population’s use of the visual modality is challenging because our linguistics toolkit is designed mainly to handle the auditory modality, putting such visual phenomena in a blind spot. Here I present a method that approximates the results of field elicitation for determining standards of form and productivity for spoken languages, but for the visual modality. The basis of this method is to substitute spoken or written elicitation sentences with video clips. Motivated by my field experiences in which I observed consultants scrutinizing the video images for meaningful communicative body behavior, I predicted that they would be sensitive to manipulations of visual practices in video clips. In this format, it is possible to apply the same principles of traditional controlled elicitation such as contrasting acceptable forms with unacceptable ones or generating new forms to test productivity.

**3.1. DESCRIPTION OF THE METHOD.** In collaboration with consultant Lourdes Brazão, I created a series of video clips in which she spoke short utterances in Nheengatú based on examples selected from the natural speech corpus, in some cases partially modified according to the goals of the exercise. I recorded the responses of nine speakers to the stimuli. In order to learn the procedure, the participants were asked to first repeat after Lourdes in two ‘practice’ clips that showed no visual time reference. Then they saw a series of clips divided into three sections. First, they saw twenty-four clips showing different combinations of spoken and visual bodily elements; twelve of these repeated the composite utterances faithfully with respect to the natural speech examples they were modeled on, and a further twelve presented versions of these examples that were modified in different ways that will be described below. This series was presented in two different random orders, mixed in with four filler clips with no celestial pointing. Second,

participants were asked to consider a series of five clips in the context of real-world scenarios in order to investigate visual time reference's status with respect to truth conditions. Third, participants evaluated a series of seven clips contrasting different clause types and polarities to test productivity beyond the composite utterance types observed in the natural speech corpus. The stimuli are summarized in Table 3.

Practice series	2 clips without visual time reference
Series 1: Standards of form	12 unmodified, 12 modified, 4 fillers
Series 2: Truth conditions	5 clips with scenarios
Series 3: Productivity	7 clips with different clause types/polarities
Total clips	36 (+ 4 fillers + 2 practice clips = 42)

TABLE 3. Breakdown of elicitation clips.

Speakers were asked to do the following for each stimulus:

- to repeat the utterance ('Say it like the speaker in the video.')
- to judge whether they considered it acceptable or unacceptable
- to provide a Portuguese translation including spoken time reference, if possible

Importantly, participants were not given any instructions mentioning gesture or the hands, and were only told to repeat and judge what Lourdes 'said' in the video clips (translated with the verbs *falar* in Portuguese and *-nheé* in Nheengatú). Even so, when participants repeated the utterance they overwhelmingly included both the auditory and the visual elements from the clip, as is discussed below.

I recruited the nine participants in the town of São Gabriel da Cachoeira by following social networks of friends and relatives of the same families with whom I had created the natural speech corpus. Elicitation sessions were filmed and lasted about one half hour each. Compared to the usual paradigm for eliciting spoken language, it was necessary to take much more care with the spatial arrangements because participants needed to be able to orient themselves accurately to the east-west axis of the sun's path. When the stimuli were recorded the speaker was facing due north, with the added clue that she was situated directly in front of the Rio Negro, a universally known landmark that is used for geocentric reference in the region. I sat to one side of the participant, facing the same direction. Then during elicitation sessions participants were seated facing due south in front of the screen. Participants were quick to equate the elicitation setting's deictic 'field' (Bühler 1982 [1934], Hanks 2005) with the space represented in the recording, and were able to interpret vectors in the recording as corresponding to vectors in the real world.

**3.2. REPETITION OF VISUAL ELEMENTS.** One of the most striking aspects of applying this elicitation paradigm designed for looking at the auditory and visual modalities together is the participants' orientation to the whole composite utterance as the repeatable unit of what was 'said'. When asked to 'say' what they had seen in the stimuli, participants almost always repeated both the auditory and visual elements together. For stimuli of unmodified versions of natural speech examples, speakers on average faithfully repeated the visual time reference 83% of the time.<sup>15</sup> Responses were judged to include

<sup>15</sup> While 83% is a high repetition rate, the 17% of cases in which speakers did not repeat the visual element of the stimulus should also be accounted for. This was a very conservative measure and excludes a number of ambiguous cases (e.g. five responses in which the time reference aligned with the Portuguese translation, not the Nheengatú repetition, and five responses in which some visual bodily behavior occurred but could not be classified as a time reference). Including these cases would bring the repetition rate to 94%. In addition, one participant patterned differently from the others, with a markedly low repetition rate (a woman currently in

a faithful repetition when a visual time reference occurred that targeted the same vector as the speaker in the clip. This was a conservative measure, permitting only a few degrees of variation; cases with any ambiguity at all were not classified as faithful repetitions. I am not aware of any studies measuring how ‘repeatable’ gesture is with respect to the speech it cooccurs with, but it seems that under similar conditions many kinds of gesture would not be repeated so frequently, so this can probably be taken as a high rate of repetition.

Two examples of faithful repetition are illustrated below; since speakers were facing the screen, here they appear to mirror the clips they responded to. In the responses to stimulus ST5, all speakers pointed to around 7 AM or 8 AM (Figure 5), while for stimulus TC3, all speakers pointed to around 3 PM or 4 PM (Figure 6). The original stimuli are shown in Figure 4.



FIGURE 4. Stimuli ST5 and TC3, mirroring the images of the responses in Figs. 5 and 6.



FIGURE 5. Nine participants’ similar responses to stimulus ST5 (8:00 ± 1hr).

As stated above, participants were given no instructions mentioning the visual modality, so the fact that they treated the visual elements in the stimuli as part of what was ‘said’ and thus part of its faithful repetition is evidence for how they perceive the auditory and visual modalities as integrated. The images in Fig. 5 and Fig. 6 also reflect what appears to be a high level of uniformity in understanding, as all participants targeted the same vector within a few degrees of variation. If we compare the rate of faith-

---

her forties who has used mainly Portuguese since her twenties; excluding her would bring the rate to 90%). Designing and implementing the stimuli were complicated, so some lack of repetition may be due to participants’ confusion with the task. Despite these factors, however, participants displayed a remarkable rate of attention to and repetition of elements in the visual modality.



FIGURE 6. Nine participants' similar responses to stimulus TC3 (16:00 ± 1hr).

ful repetitions of the stimuli set including identical spoken elements but with visual elements that were modified in ways that departed from the natural speech examples, the rate of faithful repetition decreased to 34%. It might perhaps be expected that there would be no faithful repetition of modified stimuli at all, if these were not interpretable as successful time references, but in some cases speakers were able to resort to alternative interpretations of pointing for spatial reference, as evidenced in their translations, which added spatial but not temporal information. In any case, the faithful repetition rate dropped dramatically for modified stimuli. Compared to speakers' consistent treatment of both the auditory and visual elements of the unmodified stimuli as single integrated utterances, participants oriented to the visual elements in the modified stimuli to a much lesser degree, as summarized in Table 4.

PARTICIPANT	1	2	3	4	5	6	7	8	9	AVG
% repeated unmodified stimuli	91%	91%	91%	89%	100%	100%	27%	91%	73%	83%
% repeated modified stimuli	55%	27%	18%	36%	64%	27%	0%	36%	45%	34%

TABLE 4. Relative rates of faithful repetition for unmodified vs. modified stimuli (eleven pairs of clips).<sup>16</sup>

The difference between the repetition rates of visual elements for unmodified versus modified stimuli was statistically significant in a mixed-effects logistic regression model with repetition as response, unmodified versus modified as fixed predictor, and participant and stimulus as random factors ( $\beta = -3.69$ ,  $z = -3.89$ ,  $p < 0.0001$ ). There was no significant interaction of modification with participant, indicating that participant responses to the manipulation were generally uniform. There was some degree of interaction with stimulus ( $\chi^2(2) = 15.56$ ,  $p < 0.0005$ ), indicating that repetition as a function of modification varied to some degree across stimuli. However, it should be noted that, for all pairs of stimuli, the modified clip always has lower rates of repetition than the unmodified, except in one case when it was equal.

<sup>16</sup> For this measure, I excluded one of the twelve pairs of clips in the set considered in Table 4, since those clips dealt with nighttime reference, meaning that the unmodified clip featured no visual time reference and thus no possibility of repetition. However, since the modified clip presented a visual daytime reference together with a spoken nighttime reference, it was possible for speakers to correct this conflict, meaning that it was possible to include this stimulus in the figures in Table 5 below. This difference in the data set considered in each table accounts for the small difference in some of the percentages between the two tables.

These results show that Nheengatú speakers are sensitive to manipulations of standards of form in their visual time-reference system. Testing these kinds of composite utterances for such standards is not as straightforward as the tried-and-true field method of eliciting grammaticality judgments for spoken constructions to identify grammatical and ‘starred’ examples. When asked to evaluate the stimuli metalinguistically, participants rarely went so far as to directly state that they were incorrect, instead often appealing to local dialectal differences (‘Maybe that’s how people speak where she is from’). A few participants did apply the word *errado* (‘incorrect’ in Portuguese) to the modified stimuli, but this was not consistent. When traditional metalinguistic elicitation meets its limits in this way, one possibility is to use repetition rates of stimuli as a proxy measure for grammaticality, since speakers may be less likely to repeat utterances that violate their standards of form than utterances that comply with them. Clark and Carpenter (1989) used a similar measure to study children’s usage of prepositions in oblique noun phrases, since in that case the young age of participants (rather than modality issues) made standard elicitation questions more difficult. As I found for Nheengatú, they found that faithful repetition rates decreased in the case of stimuli that departed from English’s standards of form (1989:355). They also found that not only did speakers not repeat modified stimuli, but in many cases they ‘corrected’ the modified stimuli when repeating (1989:357); the next section expands on the topic of correction.

**3.3. CORRECTION OF MODIFIED TIME REFERENCE.** In Table 4 above, faithful repetition is contrasted with all other types of response. This section considers the different ways that participants responded to the instructions to repeat what was ‘said’ in the stimulus when they did not faithfully repeat the visual elements. When the participant did nothing notable in the visual channel, or did something that could not be unambiguously judged to be a time reference, this was counted as a case without any repetition. However, there was another way that participants could respond to the modified stimuli: speakers could ‘correct’ the visual elements by altering them to conform to standards more like those observed in natural speech. For example, if the modified stimulus included a vector outside of the east-west axis, the response was judged to be a correction if the participant made a time reference along the sun’s path instead of faithfully repeating the vector seen in the stimulus. As with faithful repetition, correction was also judged conservatively, and any ambiguous cases were excluded.

PARTICIPANT	1	2	3	4	5	6	7	8	9	AVG
% repeated modified stimuli	50%	25%	17%	33%	58%	25%	0%	33%	42%	30%
% corrected modified stimuli	33%	25%	42%	33%	33%	33%	17%	50%	25%	32%
% not repeated modified stimuli	17%	50%	41%	34%	9%	42%	83%	17%	33%	38%

TABLE 5. Responses to modified stimuli (twelve single clips).

Table 5 breaks down the different types of responses speakers showed to modified stimuli. In some cases participants repeated the visual elements as seen in the stimuli. In some cases they did nothing salient, but in 32% of cases, without any special instructions to do so, they corrected the visual time reference.<sup>17</sup> The fact that this occurred

<sup>17</sup> There is significant variability in the rate of correction across participants and stimuli, as confirmed by a mixed-effects logistic regression model with response type (correction versus no repetition) as the dependent variable, with participants and stimulus as random factors. A model with both item and participant as random factors outperformed reduced models with either item or participant as the only random factor ( $\chi^2(2) = p < 0.0001$ ). This indicates that while participants were relatively uniform in their reduction of faithful repetitions as discussed in §3.2, they were less uniform in choosing an alternative to faithful repetition. Rates of correction and nonrepetition also varied to some extent depending on particular stimuli.

about a third of the time can be taken as more evidence that Nheengatú speakers orient to specific standards of form. A few examples will help to illustrate the nature of these corrections.

In both stimuli GR6A (unmodified) and GR6B (modified), the speaker said ‘They went up (the river) and came back as the day disappeared’, but in the first she made an arc from east to west, while in the second she only pointed upward (see Figure 7). Since ‘the day disappeared’ (*ara ukāyemu*) is a common way of talking about nightfall, pointing to ‘noon’ was predicted to cause a semantic conflict in the modified stimulus. In line with the results discussed in §3.3, for the unmodified stimulus all nine participants produced a time reference sweeping east-west through the arc they had seen in the video. For the modified stimulus, by contrast, eight of nine participants also swept their hand from east to west, even though they had seen a different time reference in the clip.



FIGURE 7. Unmodified and modified stimuli: GR6A (left) with appropriate time for ‘the day disappeared’ and GR6B (right) with a conflicting time reference.

- (12) Aé-ta ta-yupiri ta-uri até ara u-kāyemu.  
 3-PL 3PL-go.up 3PL-return until day 3SG-disappear  
 ‘They went up and returned until the day disappeared.’

The responses to GR7A and GR7B (Figure 8) show another instance of the same type of effect. The spoken element of the stimuli was the phrase ‘In the morning (s)he waited until noon’. In the unmodified stimulus the visual time reference was from the horizon until the sun’s zenith. In the modified stimulus the speaker pointed at the horizon but held her arm in position instead of raising it. In seven of nine cases participants faithfully repeated the visual/bodily elements of the unmodified stimulus, and in nine of nine cases they corrected the modified stimulus.



FIGURE 8. Well- and ill-formed stimuli: GR7A (left) with appropriate duration for ‘morning until noon’ and GR7B (right) with an inappropriate time.

- (13) Kuémete u-saarú =paá até yandara.  
 morning 3SG-wait =REP until noon  
 ‘In the morning (s)he waited, they say, until noon.’

Some of the modified stimuli were similar to these two examples, with the speaker in the clip making a time reference that conflicted semantically with the spoken elements.

Some clips showed pointing outside of the east-west axis altogether; to these, participants sometimes responded by correcting the vector, and other times resorted to spatial, not temporal, interpretations in order to make them intelligible. A couple of stimuli were also designed to examine the temporal alignment of visual time reference to language, with the modified time references occurring with other constituents instead of with the predicate or predicate modifiers.<sup>18</sup> These types of stimuli were difficult to design, but in many cases speakers did correct the modified version by temporally re-aligning the visual and auditory channels so it resembled the unmodified version (seven of nine participants did so for stimulus GR8B, for example). As noted in §2.4 above, the visual modality allows for time reference to precede predicates, aligning with predicate modifiers, or overlap with predicates, and in these limited results speakers prefer a close alignment of the visual time reference and the predicate.

The application of this adapted paradigm for testing standards of form through participants' repetition was not always easy, as moving from spoken to visual elicitation ventures into somewhat uncharted waters. However, even these relatively exploratory results reveal some strong and provocative trends. Participants' sensitivity to semantic conflict, spatial orientation, and temporal alignment seen in the difference in responses to the modified versus unmodified stimuli provide compelling evidence that Nheengatú speakers share similar standards of form for visual time reference. It would seem fair, then, to consider the ways in which such standards are similar to those that linguists usually illustrate for spoken elements by contrasting 'starred' and 'unstarred' examples.

**3.4. UNIFORMITY OF TEMPORAL MEANING.** In addition to repeating the utterance in the stimulus, participants were asked to provide Portuguese translations when possible. This was not always practical for a number of reasons, such as some participants' limited command of Portuguese, but many participants did provide translations, and these included equivalent spoken Portuguese times for the Nheengatú time references. This section considers the uniformity of participants' Portuguese time reference as an additional measure of the stable, conventional meanings they understood from viewing the stimuli. For example, for stimulus TC2, six of the seven participants who provided a judgment said the time reference was 7:00, while one said 8:00. The other stimuli patterned similarly, usually with one Portuguese time receiving a majority consensus. There was some variation, but responses never differed by more than one hour earlier or later than the majority consensus for any of the well-formed stimuli (see Table 6).

When the time references of participants differed, the variation was not random but rather was either one hour earlier or one hour later than the majority consensus, revealing a high degree of precision in participants' interpretations (72% exact agreement across all eleven stimuli, 100% agreement within a two-hour margin). For the modified stimuli, by contrast, in most cases participants did not provide equivalent Portuguese times, since the time references in these were intentionally unclear. For the two modified cases in which some participants did attempt to provide responses, there was no

<sup>18</sup> The full breakdown of the types of modifications in the twelve pairs of stimuli is as follows: in four clips the speaker made time references that conflicted semantically with the spoken material; in four clips the speaker pointed outside of the east-west axis used for time reference; in one pair of clips (i) a visual daytime reference was combined with a spoken nighttime reference, and (ii) a spoken time reference normally requiring a visual time reference was accompanied by no manual articulation; in two clips the time reference was moved to occur with elements other than predicates and predicate modifiers. It would of course have been even more informative to include many more clips in order to isolate each of these specific types of departures from standards of form, but this was not possible given the limited time of the field trip.

STIMULUS (# PARTICIPANTS)	< EARLIEST	MAJORITY CONSENSUS	LATEST >
TC1 (9)		12:00 100% (9/9)	
TC2 (7)		7:00 86% (6/7)	8:00 14% (1/7)
TC3 (6)	15:00 33% (2/6)	16:00 33% (2/6)	17:00 33% (2/6)
TC4 (7)	7:00 29% (2/7)	8:00 42% (3/7)	9:00 29% (2/7)
TC5 (7)		16:00 56% (4/7)	17:00 44% (3/7)
ST2 (4)	16:00 25% (1/4)	17:00 75% (3/4)	
ST4 (7)	11:00 14% (1/7)	12:00 6% (6/7)	
ST5 (7)		8:00 71% (5/7)	9:00 29% (2/7)
ST6 (6)	11:00 17% (1/6)	12:00 83% (5/6)	
GR8A (9)	11:00 12% (2/9)	12:00 88% (7/9)	
GR11A (6)		7:00 67% (4/6)	8:00 33% (2/6)

TABLE 6. Interparticipant agreement on Portuguese equivalents of Nheengatú time reference (subset of stimuli with enough Portuguese translations for comparison).

agreement among them. Compared to unmodified stimulus GR11A's agreement rate, in which four participants said 7:00 and two said 8:00, for the corresponding modified stimulus GR11B one participant said 8:00, one said 15:00, and the other four were spread out over the intervening seven-hour period. In another case, for unmodified stimulus GR8A seven participants said 12:00 and two said 11:00, while for the corresponding modified stimulus GR8B only three participants ventured guesses, one saying 7:00, one 8:00, and one 14:00, again showing variation over a seven hour period, as compared to the minimal variation in understanding with the well-formed variants.

These data show that the unmodified time references are consistently understood with the same meaning across many speakers, while the few responses to the modified stimuli show no such consistency. Of course, the meaning of a time reference in Nheengatú is not categorical in the same way that Portuguese times are, and participants' agreement is based on their shared ability to systematically translate the more analog semantics of the Nheengatú constructions into Portuguese, which forces them to choose among the quantized units of the Portuguese system. While the video paradigm is novel, this is essentially the same methodology used in all translation elicitation in which linguists can draw conclusions about semantics based on speakers' translations into a contact language.

**3.5. TRUTH CONDITIONALITY.** The visual modality may not usually be understood to encode the types of truth-conditional meanings that speakers can be held socially accountable for, except of course in the case of sign languages. The spatial vector of a deictic gesture may also be accountable in this way ('You said "here" not "here"!'), but the information at issue would depend on the specific surroundings. Nheengatú time reference, by contrast, encodes information similar to that of a spoken adverbial phrase and can be used detached from any specific context. To see if speakers considered it to have the same kind of truth conditionality as a spoken adverb, I showed the participants a series of five statements including a visual time reference, and then offered them scenarios in which everything about the statement was consistent with reality except for the time reference. For example, stimulus TC3 includes the spoken statement 'We left here and arrived here', with two celestial points providing the starting and ending moments of the duration, between 12:00 and about 15:00. When asked if the statement would still be true if the speaker had arrived in the morning, one participant pointed out that 'She **said** that she would arrive in the afternoon' ('Ela falou que ia chegar até a tarde'). See Figure 9.



FIGURE 9. Response of participant 2 to stimulus TC3: 'She said that she would arrive in the afternoon.'

For scenarios in which all elements except for the time reference were consistent with reality, speakers consistently judged the stimuli as false. To express this they applied Portuguese words such as *mentir* 'to lie', *enganhar* 'to fool', or *errar* 'to be mistaken'. The responses to stimulus TC1 provide a good illustration of such judgments. In the clip the speaker says 'Children went to bathe until right here', indicating a period from around 7:00 to 11:00. I asked if the statement would still be true if the children had arrived at 17:00, and this was one participant's response: 'No, because, she said (they left at) seven right? They came out at eleven' ('Não, pelo que, ela falou sete horas mesmo, ne? Saíram só onze horas').

As in this response, participants often used the Portuguese verb *falar* to refer to what was 'said' in reference to the combined meaning of auditory and visual elements together. In a few cases participants also used the verb 'to show' (*mostrar*) in reference to the visual elements of the stimuli, but similarly considered them to convey propositional meaning. Another participant responded as in 14 to the same stimulus, TC1, with respect to the scenario in which the children arrive at 17:00.

- (14) Não, porque ela mostrou aqui. [points up] Quer dizer que as crianças foram para tomar banho vão chegar até ela mostrou quando, quer dizer sol está no meio, quando são meiodia mais o menos. Não tem cinco horas não.

'No, because she showed (the time) here. [points up] That is to say that the children went to bathe, they will arrive when she showed (the time) that is, when the sun is in the middle, when it is more or less noon. That's not five o'clock.'

All of the participants shared this attitude and said that they would hold speakers accountable for visual temporal reference in the same way they would for any spoken adverbial element. This adds to the evidence that Nheengatú time reference productively contributes meaning similarly to a spoken adverbial modifier, a point that the next section follows up.

**3.6. MULTIMODAL PRODUCTIVITY.** The meanings of the more depictive and rhythmic kinds of visual bodily articulation cannot be separated from the auditory material they cooccur with and retain their meaningfulness, and as such do not represent tokens of types that productively add meaning to linguistic constructions. This section demonstrates that the same is not true for Nheengatú time reference. Most of the natural speech examples of the phenomenon that I collected were narrative and procedural accounts, including mainly declarative constructions about past events. In order to determine if the time references were limited to those specific speech contexts or construction types, the final part of the stimulus set asked participants to evaluate seven clips that combined temporal reference with different sentence types as well as with negative polarity and future time reference. First, my consultant Lourdes assisted me in constructing sentences that she found interpretable, and then participants judged whether they were well formed and

translatable to Portuguese with the intended reading. The following are some examples of the stimuli, showing an interrogative in 15, a negative imperative in 16, and a future reference in 17.

- (15) Indé re-sika será iké =ramé? [17:00]  
 2 2SG-arrive Q here =when  
 ‘Did you arrive at 5 PM?’
- (16) Ti re-sú yandara =ramé. [12:00]  
 NEG 2SG-go noon =when  
 ‘Do not go at 12 PM.’
- (17) Wirandé ya-sú ya-purakí iké =ramé. [8:00]  
 tomorrow 1PL-go 1PL-work here =when  
 ‘Tomorrow we will go work at 8 AM.’

For all of the stimuli, participants consistently gave the predicted reading immediately, without any apparent hesitation or confusion. From this we can conclude that the basic meaning of Nheengatú time reference is not determined by the speech it cooccurs with, but that it can be isolated, detached, and applied to other linguistic constructions of many types with a uniform, consistent contribution of meaning that can be added to any verb phrase for which time of day is relevant.

**4. DISCUSSION.** Considering the data both from the natural speech corpus and from controlled elicitation with video stimuli, we are now in a position to ask exactly what type of phenomenon Nheenagtú time reference is. Certainly it is a much more highly conventionalized practice than most well-known types of speech-accompanying gesture. If a system like Nheengatú time reference was discovered in the practices of deaf signers, it would most likely be considered part of a linguistic system and not a ‘gestural’ element. Arguments for treating Nheengatú time reference as somehow ‘nonlinguistic’ seem to risk also classifying sign languages as ‘nonlinguistic’. In fact, a number of sign languages have similar systems, like American Sign Language (ASL), which also locates different times of day along an arc modeled after the sun’s path (Baker-Shenk & Cokely 1980:183–84); ASL uses the intrinsic left/right axis, but at least one sign language, Kata Kolok, uses direct celestial pointing like Nheengatú (mentioned above; de Vos 2012, 2015). If we analyze Nheengatú time reference with the same criteria that are applied to sign languages, we find a conventionalized visual form-meaning pairing that seems to satisfy all of the criteria laid out in §1.3 for linguistic elements.

This result leaves us with two alternative analyses. In the first, Nheengatú time reference is an independent system similar to a sign language, although specialized exclusively for time of day. When speakers use this system alongside their spoken linguistic system, temporal meaning is added at the pragmatic level. In the second analysis, Nheengatú time reference is simply another type of adverb in a single hybrid, multimodal linguistic system. In this case the temporal meaning would be added at the morphosyntactic level, like with any other adverb. We can revisit a previous example to think about the respective implications of these two analyses. In 18 Marcilia says the words ‘they dance’ and at the same time makes a sweeping time reference covering the span of most of a day.

- (18) Ta-purasi.  
 3PL-dance  
 ‘They dance from 10 AM to 6 PM.’



In the pragmatic analysis, both channels would be encoded separately; this necessitates proposing some mechanism for combining them at the discourse level. The morphosyntactic analysis is simpler in that it does not require such a mechanism, but the trade-off is that the hierarchical morphosyntactic structures posited become more complex (see Figure 10).

a. Pragmatic analysis:

VISUAL CHANNEL:	[Adv 10 AM–6 PM]	} ‘They dance from 10 AM to 6 PM.’
AUDITORY CHANNEL:	[S [NP 3PL] [VP [V dance]]]	

b. Morphosyntactic analysis:

HYBRID CHANNEL:	[S [NP 3PL] [VP [Adv 10AM–6PM] [V dance]]] → ‘They dance from 10 AM to 6 PM.’
-----------------	---

FIGURE 10. Two different perspectives on visual time reference in Nheengatú.

These two alternatives could have quite different implications depending on linguists’ different theoretical approaches; the consequence of choosing the morphosyntactic analysis would be that whatever terms or concepts are applied to the auditory channel would have to accommodate elements of the visual channel as well. The pragmatic analysis avoids facing this challenge head on. Both analyses support the position that communicative moves integrate multiple modalities and so are best approached as a whole (Enfield 2009; also Clark 1996, Bavelas & Chovil 2000, Goodwin 2000; Kendon takes a similar position regarding a ‘single plan of action’ across modalities (1997:110–11); McNeill 1992 applies the ‘single plan’ principle to speech production). But the morphosyntactic analysis goes further, aligning with a number of scholars who have advocated more controversial proposals for thinking in terms of ‘mixed’ (Slama-Cazacu 1976) or ‘multichannel’ (Jouitteau 2004a, 2007) syntax or ‘multimodal grammar’ (Fricke 2012).

On the one hand, languages all around the world have resources for communicating information about time of day, and these are usually treated as adverbial modifiers within verb phrases, so it is unclear why Nheengatú should be treated differently simply on the basis of modality. As sign language research has made abundantly clear (Klima & Bellugi 1979, Liddell 2002, 2003, among many others), conventionalized linguistic elements cannot be identified based on their modality of expression, but only by their semiotic properties (Okrent 2002), and the data presented in §2 and §3 illustrate how Nheengatú time reference satisfies all of the major proposed criteria for linguistic elements. Excluding it from morphosyntactic analysis would seem to be based more on its modality than on any other criterion. There is also good evidence that the speakers themselves consider both channels together to be part of a well-formed utterance since they consistently repeat bimodally when asked only to repeat what was ‘said’ in the stimuli.

On the other hand, the pragmatic analysis requires simpler morphosyntactic structures, and avoids reworking traditional linguistic concepts that were not designed for the visual modality. Keeping the two modalities separate necessitates proposing some extra integrating mechanism at the pragmatic level, but this is not unprecedented. The field of gesture studies has documented a wide range of visual elements that add information to utterances in ways that do not necessarily make them morphosyntactic constituents with respect to the spoken language.

However, it is clear that Nheengatú time reference does not have the idiosyncratic and context-dependent characteristics more commonly associated with the visual channel when used together with spoken language. Speakers showed overwhelming agreement on the meaning of visual time references and were able to easily and consistently parse new constructions with context-independent meanings added by the visual channel. Yet while Nheengatú visual time reference is quite language-like on its own terms, it may be too early to argue that it is necessarily part of a modally hybrid grammar, rather than just a system that is used in parallel with the spoken morphosyntax. Even after decades of discussion, linguists still disagree on similar issues such as the question of whether intonation systems are best treated as paralinguistic or linguistic (Ladd 2008:3–42, Kohler 2012, etc.), so it may be premature to hope to determine which analysis best accounts for the Nheengatú data before we know more about similar phenomena crosslinguistically.

The reports of visual time-reference systems similar to that of Nheengatú from around the world cited in §2.1 suggest that such phenomena are still to be fully described for many languages and speech communities. Further diverse kinds of conventionalized and semi-conventionalized combinations of auditory and visual elements have been described for a number of languages (Jouitteau 2004a,b, 2007 for French; Slama-Cazacu 1976 for Romanian; Fricke 2012 for German; Le Guen 2011 for Yucatec Maya; Enfield 2009 for Lao; Wilkins 2003 for Arrernte, among others).<sup>19</sup> The fact that these types of phenomena turn up suggests that it is not strictly true that gesture will develop conventionalized, language-like properties only when the spoken channel is not available. This is probably better framed as a tendency rather than an absolute, as a gradient difference between spoken and signed languages rather than a binary one. Practical considerations for keeping hands or gaze free for other purposes may offer some explanation for why such practices are not more common, along with other factors including literacy and the use of speech-only technologies like telephones. But these do not seem to constitute firm constraints on their development.

Applying variants of the video stimulus-based methodology tests that I developed for Nheengatú to other languages and multimodal practices could help reveal what types of similar processes occur crosslinguistically. It may be that the lack of attestation of more such practices in the literature is not because they do not exist but because they have yet to be studied. The fact that during my first weeks of fieldwork Nheengatú temporal gestures went unperceived until my consultant corrected my spoken-language bias illustrates how easy it could be to miss them, especially when working without video. Descriptive grammars rarely address multimodality, so while linguistic typology has been able to advance by drawing on hundreds of existing accounts of spoken languages, few such accounts exist for multimodal practices, and as a result our knowledge of multimodality crosslinguistically is almost nothing by comparison. Thankfully, since most recent language documentation and description efforts include the collection of video data, grammar writers may be able to begin including accounts of visual bodily practices in their descriptions, making future typological work possible. Once these new

<sup>19</sup> In his review of Kendon 2004, Wilkins takes the following position on the potential for analyzing an even wider range of ‘gestural’ elements than I do here more in terms of linguistic structure and linguistic systems: ‘Depictive imagistic “gesticulations” that can serve a referential function may not be digital or analytic or simply concatenative and segmentable in the sense of words and morphemes, but their analog and suprasegmental or synthetic nature does not make them any less subject to convention, and does not deny them combinatorial constraints or rules of structural form’ (2006:131–32).

data are accounted for, I suspect that modal hybridity may turn out to be more common than we presently know.

## REFERENCES

- AIKHENVALD, ALEXANDRA Y. 2002. *Language contact in Amazonia*. Oxford: Oxford University Press.
- AIKHENVALD, ALEXANDRA Y. 2003. Multilingualism and ethnic stereotypes: The Tariana of northwest Amazonia. *Language in Society* 32.1–21. DOI: 10.1017/S0047404503321013.
- BAKER-SHENK, CHARLOTTE, and DENNIS COKELY. 1980. *American Sign Language: A teacher's resource text on grammar and culture*. Silver Spring, MD: T. J. Publishers.
- BAVELAS, JANET B., and NICOLE CHOUIL. 2000. Visible acts of meaning: An integrated message model of language use in face-to-face dialogue. *Journal of Language and Social Psychology* 19.163–94. DOI: 10.1177/0261927X00019002001.
- BOHNEMEYER, JÜRGEN. 2009. Temporal anaphora in a tenseless language. *The expression of time in language*, ed. by Wolfgang Klein and Ping Li, 83–128. Berlin: Mouton de Gruyter.
- BOHNEMEYER, JÜRGEN. 2011. Spatial frames of reference in Yucatec: Referential promiscuity and task-specificity. *Language Sciences* 33.892–914. DOI: 10.1016/j.langsci.2011.06.009.
- BORODITSKY, LERA, and ALICE GABY. 2010. Remembrances of times east: Absolute spatial representations of time in an Australian Aboriginal community. *Psychological Science* 21.11.1635–39. DOI: 10.1177/0956797610386621.
- BROOKES, HEATHER J. 2001. *O clever 'He's streetwise.'* When gestures become quotable: The case of the *clever* gesture. *Gesture* 1.2.167–84. DOI: 10.1075/gest.1.2.05bro.
- BROOKES, HEATHER J. 2004. A repertoire of South African quotable gestures. *Journal of Linguistic Anthropology* 14.2.186–224. DOI: 10.1525/jlin.2004.14.2.186.
- BROOKES, HEATHER J. 2005. What gestures do: Some communicative functions of quotable gestures in conversations among Black urban South Africans. *Journal of Pragmatics* 32.2044–85. DOI: 10.1016/j.pragma.2005.03.006.
- BROOKES, HEATHER J. 2011. *Amangama amathathu* 'The three letters': The emergence of a quotable gesture (emblem). *Gesture* 11.2.194–218. DOI: 10.1075/gest.11.2.05bro.
- BÜHLER, KARL. 1982 [1934]. The deictic field of language and deictic words. *Speech, place and action*, ed. by Robert J. Jarvella and Wolfgang Klein, 9–30. New York: Wiley.
- CLARK, EVE V., and KATHIE L. CARPENTER. 1989. On children's uses of *from*, *by* and *with* in oblique noun phrases. *Journal of Child Language* 16.349–64. DOI: 10.1017/S030500090001045X.
- CLARK, HERBERT H. 1996. *Using language*. Cambridge: Cambridge University Press.
- CLARK, WILLIAM P. 1885. *The Indian sign language*. Philadelphia: Hamersly.
- COOPERRIDER, KENSY, and RAFAEL NÚÑEZ. 2009. Across time, across the body: Transversal temporal gestures. *Gesture* 9.2.181–206. DOI: 10.1075/gest.9.2.02coo.
- COOPERRIDER, KENSY, and RAFAEL NÚÑEZ. 2012. Nose-pointing: Notes on a facial gesture of Papua New Guinea. *Gesture* 12.2.103–30. DOI: 10.1075/gest.12.2.01coo.
- DA CRUZ, ALINE. 2011. *Fonologia e gramática do Nheengatú: A língua geral falada pelos povos Baré, Warekena e Baniwa*. (LOT dissertation series 280.) Utrecht: LOT.
- DAVIS, JEFFREY E. 2010. *Hand talk: Sign language among American Indian nations*. Cambridge: Cambridge University Press.
- DE VOS, CONNIE. 2012. *Sign-spatiality in Kata Kolok: How a village sign language in Bali inscribes its signing space*. Nijmegen: Radboud University dissertation.
- DE VOS, CONNIE. 2015. The Kata Kolok pointing system: Morphemization and syntactic integration. *Topics in Cognitive Science* 7.1.150–68. DOI: 10.1111/tops.12124.
- DELAPORTE, YVES, and EMILY SHAW. 2009. Gesture and signs through history. *Gesture* 9.1.35–60. DOI: 10.1075/gest.9.1.02sha.
- EFRON, DAVID. 1972 [1941]. *Gesture, race and culture*. The Hague: Mouton.
- EKMAN, PAUL. 1976. Movements with precise meanings. *Journal of Communication* 26.3.14–26. DOI: 10.1111/j.1460-2466.1976.tb01898.x.
- EMMOREY, KAREN; HELSA B. BORINSTEIN; and ROBIN THOMPSON. 2005. Bimodal bilingualism: Code-blending between spoken English and American Sign Language. *Proceedings of the 4th International Symposium on Bilingualism*, ed. by James Cohen, Kara T.

- McAlister, Kellie Rolstad, and Jeff MacSwan, 663–73. Somerville, MA: Cascadilla. Online: <http://www.lingref.com/isb/4/0511SB4.PDF>.
- EMMOREY, KAREN; HELSA B. BORINSTEIN; ROBIN THOMPSON; and TAMAR H. GOLLAN. 2008. Bimodal bilingualism. *Bilingualism: Language and Cognition* 11.1.43–61. DOI: 10.1017/S1366728907003203.
- ENFIELD, NICK J. 2001. ‘Lip-pointing’: A discussion of form and function with reference to data from Laos. *Gesture* 1.2.185–211. DOI: 10.1075/gest.1.2.06enf.
- ENFIELD, NICK J. 2005. The body as a cognitive artifact in kinship representations: Hand gesture diagrams by speakers of Lao. *Current Anthropology* 46.1.51–81. DOI: 10.1086/425661.
- ENFIELD, NICK J. 2009. *The anatomy of meaning: Speech, gesture, and composite utterances*. Cambridge: Cambridge University Press.
- ENFIELD, NICK J. 2013. A ‘composite utterances’ approach to meaning. *Body—language—communication: An international handbook on multimodality in human interaction*, vol. 1, ed. by Cornelia Müller, Ellen Fricke, Silvia Ladewig, Allan Cienki, David McNeill, and Sedinha Teßendorf, 689–706. Berlin: Mouton de Gruyter.
- ENFIELD, NICK J.; SOTARO KITA; and JAN-PETER DE RUITER. 2007. Primary and secondary pragmatic functions of pointing gestures. *Journal of Pragmatics* 39.1722–41. DOI: 10.1016/j.pragma.2007.03.001.
- EPPS, PATIENCE. 2005. Areal diffusion and the development of evidentiality: Evidence from Hup. *Studies in Language* 29.3.617–50. DOI: 10.1075/sl.29.3.04epp.
- EPPS, PATIENCE. 2009. Language classification, language contact, and Amazonian prehistory. *Language and Linguistics Compass* 3.2.581–606. DOI: 10.1111/j.1749-818X.2009.00126.x.
- EVERETT, DANIEL L. 2005. Cultural constraints on grammar and cognition in Pirahã: Another look at the design features of human language. *Current Anthropology* 46.4.621–46. DOI: 10.1086/431525.
- FABIAN, STEPHEN M. 1992. *Space-time of the Bororo of Brazil*. Gainesville: University of Florida Press.
- FLOYD, SIMEON. 2007. Changing times and local terms on the Rio Negro, Brazil: Amazonian ways of depolarizing epistemology, chronology and cultural change. *Latin American and Caribbean Ethnic Studies* 2.2.111–40. DOI: 10.1080/17442220701489548.
- FRANK, MICHAEL C.; DANIEL L. EVERETT; EVELINA FEDORENKO; and EDWARD GIBSON. 2008. Number as a cognitive technology: Evidence from Pirahã language and cognition. *Cognition* 108.819–24. DOI: 10.1016/j.cognition.2008.04.007.
- FRICKE, ELLEN. 2012. *Grammatik multimodal: Wie Wörter und Gesten zusammenwirken*. Berlin: De Gruyter.
- GOLDIN-MEADOW, SUSAN. 2005. Watching language grow. *Proceedings of the National Academy of Sciences* 102.7.2271–72. DOI: 10.1073/pnas.0500166102.
- GOLDIN-MEADOW, SUSAN; DAVID MCNEILL; and JENNY SINGLETON. 1996. Silence is liberating: Removing the handcuffs on grammatical expression in the manual modality. *Psychological Review* 103.1.34–55. DOI: 10.1037/0033-295X.103.1.34.
- GOLDIN-MEADOW, SUSAN; CAROLYN MYLANDER; and AMY FRANKLIN. 2007. How children make language out of gesture: Morphological structure in gesture systems developed by American and Chinese deaf children. *Cognitive Psychology* 55.87–135. DOI: 10.1016/j.cogpsych.2006.08.001.
- GOODWIN, CHARLES. 2000. Action and embodiment within situated human interaction. *Journal of Pragmatics* 32.1489–1522. DOI: 10.1016/S0378-2166(99)00096-X.
- GORDON, PETER. 2004. Numerical cognition without words: Evidence from Amazonia. *Science* 306.496–99. DOI: 10.1126/science.1094492.
- HANKS, WILLIAM F. 2005. Explorations in the deictic field. *Current Anthropology* 46.2.191–220. DOI: 10.1086/427120.
- HANNA, BARBARA E. 1996. Defining the emblem. *Semiotica* 112.3–4.289–358. DOI: 10.1515/semi.1996.112.3-4.289.
- HAVILAND, JOHN B. 1993. Anchoring, iconicity, and orientation in Guugu Yimithirr pointing gestures. *Journal of Linguistic Anthropology* 3.1.3–45. DOI: 10.1525/jlin.1993.3.1.3.
- HAVILAND, JOHN B. 2000. Pointing, gesture spaces, and mental maps. *Language and gesture: Window into thought and action*, ed. by David McNeill, 13–46. Cambridge: Cambridge University Press.

- HAVILAND, JOHN B. 2003. How to point in Zinacantán. In *Kita 2003*, 139–69.
- JACKSON, JEAN E. 1983. *The fish people: Linguistic exogamy and Tukanoan identity in north-west Amazonia*. Cambridge: Cambridge University Press.
- JOHNSTON, TREVOR. 2013. Formational and functional characteristics of pointing signs in a corpus of Auslan (Australian Sign Language): Are the data sufficient to posit a grammatical class of ‘pronouns’ in Auslan? *Corpus Linguistics and Linguistic Theory* 9.1. 109–59.
- JOUITTEAU, MÉLANIE. 2004a. Gestures as expletives, multichannel syntax. *West Coast Conference on Formal Linguistics (WCCFL)* 23.422–35.
- JOUITTEAU, MÉLANIE. 2004b. Gesturing at the interface. [*Domain [E] s*], *actes du colloque JEL’2004*, ed. by Olivier Crouzet, Hamida Demirdache, and Sophie Wauquières-Gravelines, 183–88. Nantes: Université de Nantes/Naoned.
- JOUITTEAU, MÉLANIE. 2007. Listen to the sound of salience: Multichannel syntax of Q particles. *Romance languages and linguistic theory 2005: Selected papers from ‘Going Romance’, Utrecht, 8–10 December 2005*, ed. by Sergio Baauw, Frank Drijkoningen, and Manuela Pinto, 185–200. Amsterdam: John Benjamins.
- KEGL, JUDY; ANN SENGHAS; and MARIE COPPOLA. 1999. Creation through contact: Sign language emergence and sign language change in Nicaragua. *Comparative grammatical change: The intersection of language acquisition, creole genesis, and diachronic syntax*, ed. by Michel DeGraff, 179–237. Cambridge, MA: MIT Press.
- KENDON, ADAM. 1987. Speaking and signing simultaneously in Warlpiri sign language users. *Multilingua* 6.25–68. DOI: 10.1515/mult.1987.6.1.25.
- KENDON, ADAM. 1988a. *Sign languages of Aboriginal Australia: Cultural, semiotic and communicative perspectives*. Cambridge: Cambridge University Press.
- KENDON, ADAM. 1988b. Parallels and divergences between Warlpiri sign language and spoken Warlpiri: Analyses of signed and spoken discourses. *Oceania* 58.239–54. Online: <http://www.jstor.org/stable/40331045>.
- KENDON, ADAM. 1992. Some recent work from Italy on quotable gestures (‘emblems’). *Journal of Linguistic Anthropology* 2.1.77–93. DOI: 10.1525/jlin.1992.2.1.92.
- KENDON, ADAM. 1997. Gesture. *Annual Review of Anthropology* 26.109–28. DOI: 10.1146/annurev.anthro.26.1.109.
- KENDON, ADAM. 2004. *Gesture: Visible action as utterance*. Cambridge: Cambridge University Press.
- KENDON, ADAM. 2008. Some reflections on the relationship between ‘gesture’ and ‘sign’. *Gesture* 8.3.348–66. DOI: 10.1075/gest.8.3.05ken.
- KENDON, ADAM. 2010. Pointing and the problem of ‘gesture’: Some reflections. *Rivista di Psicolinguistica Applicata* 10.3.19–30.
- KENDON, ADAM, and LAURA VERSANTE. 2003. Pointing by hand in ‘Neapolitan’. In *Kita 2003*, 109–37.
- KITA, SOTARO (ed.) 2003. *Pointing: Where language, culture, and cognition meet*. Mahwah, NJ: Lawrence Erlbaum.
- KITA, SOTARO. 2009. Cross-cultural variation of speech-accompanying gesture: A review. *Language and Cognitive Processes* 24.2.145–67. DOI: 10.1080/01690960802586188.
- KITA, SOTARO, and JAMES ESSEGBEY. 2001. Pointing left in Ghana: How a taboo on the use of the left hand influences gestural practice. *Gesture* 1.1.73–95. DOI: 10.1075/gest.1.1.06kit.
- KLIMA, EDWARD S., and URSULA BELLUGI. 1979. *The signs of language*. Cambridge, MA: Harvard University Press.
- KOHLER, KLAUS J. 2012. From phonetics to phonology and back again. *Phonetica* 69.254–73. DOI: 10.1159/000351218.
- LADD, ROBERT D. 2008. *Intonational phonology*. 2nd edn. Cambridge: Cambridge University Press.
- LE GUEN, OLIVIER. 2011. Modes of pointing to existing spaces and the use of frames of reference. *Gesture* 11.3.271–307. DOI: 10.1075/gest.11.3.02leg.
- LE GUEN, OLIVIER, and LORENA POOL BALAM. 2012. No metaphorical timeline in gesture and cognition among Yucatec Mayas. *Frontiers in Psychology* 3:271. DOI: 10.3389/fpsyg.2012.00271.
- LEVINSON, STEPHEN C. 2003. *Space in language and cognition*. Cambridge: Cambridge University Press.

- LEVINSON, STEPHEN C., and ASIFA MAJID. 2013. The island of time: Yéfi Dnye, the language of Rossel Island. *Frontiers in Psychology* 4:61. DOI: 10.3389/fpsyg.2013.00061.
- LEVINSON, STEPHEN C., and DAVID P. WILKINS (eds.) 2006. *Grammars of space*. Cambridge: Cambridge University Press.
- LIDDELL, SCOTT K. 2000. Indicating verbs and pronouns: Pointing away from agreement. *The signs of language revisited*, ed. by Karen Emmorey and Harlan Lane, 303–20. Mahwah, NJ: Lawrence Erlbaum.
- LIDDELL, SCOTT K. 2002. Modality effects and conflicting agendas. *The study of sign languages: Essays in honor of William C. Stokoe*, ed. by David F. Armstrong, Michael A. Karchmer, and John V. Van Cleve, 53–81. Washington, DC: Gallaudet University Press.
- LIDDELL, SCOTT K. 2003. *Grammar, gesture, and meaning in American Sign Language*. Cambridge: Cambridge University Press.
- LILLO-MARTIN, DIANE. 2002. Where are all the modality effects? In Meier et al., 241–62.
- LILLO-MARTIN, DIANE; RONICE MÜLLER DE QUADROS; HELEN KOULIDOBROVA; and DEBORAH CHEN PICHLER. 2010. Bimodal bilingual cross-language influence in unexpected domains. *Language acquisition and development: Proceedings of GALA 2009*, ed. by João Costa, Maria Lobo, and Fernanda Pratas, 264–75. Newcastle: Cambridge Scholars Publishing.
- MALLERY, GARRICK. 1880. The sign language of the North American Indians. *American Annals of the Deaf* 25.1.1–20.
- MCNEILL, DAVID. 1985. So you think gestures are nonverbal? *Psychological Review* 92.3.350–71. DOI: 10.1037/0033-295X.92.3.350.
- MCNEILL, DAVID. 1992. *Hand and mind: What gestures reveal about thought*. Chicago: University of Chicago Press.
- MCNEILL, DAVID. 2000. Introduction. *Language and gesture*, ed. by David McNeill, 1–11. Cambridge: Cambridge University Press.
- MEIER, RICHARD P. 2002. Why different, why the same? Explaining effects and non-effects of modality upon linguistic structure in sign and speech. In Meier et al., 1–25.
- MEIER, RICHARD P.; KEARSY CORMIER; and DAVID QUINTO-POZOS (eds.) 2002. *Modality and structure in signed and spoken languages*. Cambridge: Cambridge University Press.
- MOORE, DENNY; SIDNEY FACUNDES; and NÁDIA PIRES. 1994. Nheengatu (Língua Geral Amazônica), its history, and the effects of language contact. *Survey of Californian and other Indian languages, vol. 8: Proceedings of the Meetings of the SSILA, July 2–4, 1993*, 93–118. Online: <http://linguistics.berkeley.edu/~survey/documents/survey-reports/survey-report-8.07-moore-et-al.pdf>.
- NILSSON, MARTIN P. 1920. *Primitive time reckoning*. London: Humphrey Milford.
- NÚÑEZ, RAFAEL E., and EVE SWEETSER. 2006. With the future behind them: Convergent evidence from Aymara language and gesture in the crosslinguistic comparison of spatial construals of time. *Cognitive Science* 30.3.401–50. DOI: 10.1207/s15516709cog0000\_62.
- OKRENT, ARIKA. 2002. A modality-free notion of gesture and how it can help us with the morpheme vs. gesture question in sign language linguistics (or at least give us some criteria to work with). In Meier et al., 167–74.
- PAYRATÓ, LLUÍS. 2008. Past, present, and future research on emblems in the Hispanic tradition: Preliminary and methodological considerations. *Gesture* 8.1.5–21. DOI: 10.1075/gest.8.1.03pay.
- PIZZUTO, ELENA ANTINORO, and MICAELA CAPOBIANCO. 2008. Is pointing ‘just’ pointing? Unraveling the complexity of indexes in spoken and signed discourse. *Gesture* 8.1.82–103. DOI: 10.1075/gest.8.1.07piz.
- REITER, SABINE. 2012. The interaction of ideophones, gestures and lexical verbs in the presentation of motion events in Awetí discourse. Paper presented at the 54th International Congress of Americanists, Vienna, July 15–20.
- RODRIGUES, ARYON D. 1986. *Línguas brasileiras: Para o conhecimento das línguas indígenas*. São Paulo: Edições Loyola.
- RODRIGUES, ARYON D. 1996. As Línguas Gerais Sul-Americanas. *Papia* 4.2.6–18. Online: <http://www.revistas.fflch.usp.br/papia/article/view/1791/1602>.
- SANDLER, WENDY. 2003. On the complementarity of signed and spoken languages. *Language competence across populations: Towards a definition of SLI*, ed. by Yonata Levy and Jeannette Schaeffer, 383–409. Mahwah, NJ: Lawrence Erlbaum.

- SANDLER, WENDY. 2009. Symbiotic symbolization by hand and mouth in sign language. *Semiotica* 174.1/4.241–75. DOI: 10.1515/semi.2009.035.
- SANDLER, WENDY, and DIANE LILLO-MARTIN. 2005. Sign language. *Contemporary linguistics: An introduction*, 5th edn., ed. by William O'Grady, John Archibald, Mark Aronoff, and Janie Rees-Miller, 343–60. Boston: Bedford St. Martins.
- SHERZER, JOEL. 1972. Verbal and nonverbal deixis: The pointed lip gesture among the San Blas Cuna. *Language in Society* 2.117–31. DOI: 10.1017/S0047404500000087.
- SHERZER, JOEL. 1991. The Brazilian thumbs-up gesture. *Journal of Linguistic Anthropology* 1.2.189–97. DOI: 10.1525/jlin.1991.1.2.189.
- SILVA, ALCIONILIO BRUZZI ALVES DA. 1962. *A civilização indígena do Uaupés*. São Paulo: Centro de Pesquisas de Iauareté.
- SLAMA-CAZACU, TATIANA. 1976. Nonverbal components in message sequence: 'Mixed syntax'. *Language and man: Anthropological issues*, ed. by William C. McCormack and Stephen A. Wurm, 217–22. The Hague: Mouton.
- SMITH, CARLOTA. 1997. *The parameter of aspect*. Dordrecht: Kluwer.
- SMITH, CARLOTA. 2008. Time with and without tense. *Time and modality*, ed. by Jacqueline Guéron and Jacqueline Lecarme, 227–49. Dordrecht: Springer.
- SORENSEN, ARTHUR P., JR. 1967. Multilingualism in the Northwest Amazon. *American Anthropologist* 69.670–84. DOI: 10.1525/aa.1967.69.6.02a00030.
- STENZEL, KRISTINE. 2005. Multilingualism in the Northwest Amazon, revisited. *Memorias del Congreso de Idiomas Indígenas de Latinoamérica-II 27–29 Oct.*, University of Texas at Austin. Online: [http://www.ailla.utexas.org/site/cilla2/Stenzel\\_CILLA2\\_vaupes.pdf](http://www.ailla.utexas.org/site/cilla2/Stenzel_CILLA2_vaupes.pdf).
- STOKOE, WILLIAM C. 1960. *Sign language structure: An outline of the visual communication systems of the American deaf*. (Studies in linguistics: Occasional papers 8.) Buffalo: University of Buffalo.
- STOKOE, WILLIAM C.; DOROTHY C. CASTERLINE; and CARL G. CRONEBERG. 1965. *Dictionary of American Sign Language on linguistic principles*. Washington, DC: Gallaudet University Press.
- SWEETSER, EVE. 2009. What does it mean to compare language and gesture? Modalities and contrasts. *Crosslinguistic approaches to the psychology of language: Studies in the tradition of Dan Isaac Slobin*, ed. by Jiansheng Guo, Elena Lieven, Nancy Budwig, Susan Ervin-Tripp, Keiko Nakamura, and Seyda Ozcaliskan, 357–66. New York: Psychology Press.
- TAYLOR, GERALD. 1985. Apontamentos sobre o neengatu falado no Rio Negro, Brasil. *Amerindia* 10.3.23.
- TOMASELLO, MICHAEL. 2008. *Origins of human communication*. Cambridge, MA: MIT Press.
- TOMKINS, WILLIAM. 1969 [1931]. *Indian sign language*. Toronto: Dover.
- TONHAUSER, JUDITH. 2011. Temporal reference in Paraguayan Guaraní, a tenseless language. *Linguistics and Philosophy* 34.3.257–303. DOI: 10.1007/s10988-011-9097-2.
- WILKINS, DAVID. 2003. Why pointing with the index finger is not a universal (in sociocultural and semiotic terms). In *Kita* 2003, 171–215.
- WILKINS, DAVID P. 2006. Review of Kendon 2004. *Gesture* 6.1.119–44. DOI: 10.1075/gest.6.1.08wil.

Wundtlaan 1  
6525 XD Nijmegen  
The Netherlands  
[simeon.floyd@mpi.nl]

[Received 22 July 2011;  
revision invited 19 March 2012;  
revision received 23 April 2013;  
revision invited 31 January 2014;  
revision received 11 April 2014;  
accepted 15 February 2015]