



Jordan Zlatev\*, Simon Devylder, Rebecca Defina, Kalina Moskaluk  
and Linea Brink Andersen

# Analyzing polysemiosis: language, gesture, and depiction in two cultural practices with sand drawing

<https://doi.org/10.1515/sem-2022-0102>

Received September 13, 2022; accepted April 20, 2023; published online June 26, 2023

**Abstract:** Human communication is by default *polysemiotic*: it involves the spontaneous combination of two or more *semiotic systems*, the most important ones being *language*, *gesture*, and *depiction*. We formulate an original cognitive-semiotic framework for the analysis of polysemiosis, contrasting this with more familiar systems based on the ambiguous term “multimodality.” To be fully explicit, we developed a coding system for the analysis of polysemiotic utterances containing speech, gesture, and drawing, and implemented this in the ELAN video annotation software. We used this to analyze 23 video-recordings of sand drawing performances on Paama, Vanuatu and 20 sand stories of the Pitjantjatjara culture in Central Australia. Methodologically we used the *conceptual-empirical loop* of cognitive semiotics: our theoretical framework guided general considerations, such as distinguishing between the “tiers” of gesture and depiction, and the three kinds of semiotic grounds (iconic, indexical, symbolic), but the precise decisions on how to operationalize these were made only after extensive work with the material. We describe the coding system in detail and provide illustrative examples from the Paamese and Pitjantjatjara data, remarking on both similarities and differences in the polysemiosis of the two cultural practices. We conclude by summarizing the contributions of the study and point to some directions for future research.

**Keywords:** cognitive semiotics; diagrams; Paama; Pitjantjatjara; sand stories; semiotic grounds; sign systems

---

\***Corresponding author: Jordan Zlatev**, Lund University, Lund, Sweden,

E-mail: [jordan.zlatev@semiotik.lu.se](mailto:jordan.zlatev@semiotik.lu.se). <https://orcid.org/0000-0002-6096-4763>

**Simon Devylder**, Lund University, Lund, Sweden; and The Arctic University of Norway, Tromsø, Norway,

E-mail: [simon.devylder@uit.no](mailto:simon.devylder@uit.no). <https://orcid.org/0000-0001-6758-9874>

**Rebecca Defina**, University of Melbourne, Parkville, VIC, Australia,

E-mail: [rebecca.defina@unimelb.edu.au](mailto:rebecca.defina@unimelb.edu.au)

**Kalina Moskaluk and Linea Brink Andersen**, Lund University, Lund, Sweden,

E-mail: [kalina.moskaluk4034@gmail.com](mailto:kalina.moskaluk4034@gmail.com) (K. Moskaluk), [lineaba@gmail.com](mailto:lineaba@gmail.com) (L.B. Andersen)

# 1 Introduction

Language, realized as speech, signed language or writing, is a universal and uniquely human semiotic system. But it is by far not the only one, since all known cultures make extensive use of (more or less) deliberate, expressive movements of the hands and the rest of the body: *gesture* (Kendon 2004). And a great majority of human cultures have adopted the production of marks on two-dimensional surfaces to represent various objects and events: *depiction*, using this for varieties of functions, from religious ritual to advertising (Sonesson 1989). In fact, human communication is by default *polysemiotic*: it involves the spontaneous combination of two or more *semiotic systems* (Louhema et al. 2019; Zlatev 2019).

But what exactly is a “semiotic system,” what kinds of semiotic systems are there, and how do people use them in different cultural practices? Further, how does polysemiosis relate to the popular, but (as we argue in Section 2) ambiguous and problematic notion of “multimodality”? In this paper we address these questions, following a relatively recent approach within the new discipline of *cognitive semiotics* (Sonesson 2007, 2010; Zlatev 2015a; Zlatev et al. 2016). This approach has been applied successfully to the study of polysemiosis in the context of language evolution (Zlatev 2019; Zlatev et al. 2020), intersemiotic translation (Diget 2019; Louhema et al. 2019), and street art (Stampoulidis 2021).

After having established some necessary conceptual ground in Section 2, we turn to the empirical side of the *conceptual-empirical loop* that is typical for cognitive semiotics (Zlatev 2015a), and introduce a coding system for polysemiosis, implemented in the ELAN software (Wittenburg et al. 2006) with distinct layers of analysis (so-called “tiers”) for the semiotic systems of speech, gesture and depiction, including operationalizations of the different kinds of signs in each system. This coding system was developed through several iterations of analyzing video-recordings that we made of polysemiotic cultural practices in two Indigenous cultures: Paamese in Vanuatu (Devylde 2019a, 2022) and Pitjantjatjara in Australia (Eickelkamp 2011; Tjitayi and Lewis 2011). These two practices differ from each other in many ways but share a key common aspect: *sand drawing*.

Vanuatu sand drawing is a cultural practice used in a few northern and central islands of Vanuatu and it has been recognized as part of the Intangible Cultural Heritage of Humanity (UNESCO 2006). It is an ancient practice, first documented by Deacon and Wedgwood (1934) on the island of Malekula, but it is currently highly endangered. Missionaries considered sand drawing to be associated with witchcraft and therefore proscribed it. Mentions of the practice, or qualitative analyses of individual patterns can be found scattered in the anthropological literature (Gell 1998; Layard 1936; Patterson 2006; Rio 2005; Taylor 2005), while systematic analyses of a

corpus of Vanuatu sand drawings are recent (Baron 2020; Devylder 2022). Crucially for our purposes, during or after drawing, the performer also gestures and speaks in order to help the addressee understand the significance of the practice.

Pitjantjatjara sand storytelling is an ancient and living cultural practice among women and girls in Central and Western Australia (Eickelkamp 2011; Munn 1973).<sup>1</sup> It can be a group activity with people exchanging stories, each drawing in their own spaces, or with one person acting as the storyteller while the others listen and sometimes comment. It can also be a private activity performed solitarily or with a friend. The types of stories told cover a wide range including myths, folktales, personal histories, accounts of past events, gossip, discussion or commentary of issues, and plans or dreams for the future. The interaction between drawing, speaking, and gesturing is conspicuous in the practice.

For the purpose of the study, we performed systematic analysis of 23 video-recordings of sand drawing performances on Paama, Vanuatu and 20 sand stories in Central Australia. We present examples from this data to explain the coding system in Section 3, after defining our main theoretical concepts, and relate our framework briefly to other theoretical notions in Section 2, as mentioned above. In Section 4, we explain how the system was used for the analysis of the data, involving some necessary operational definitions. Sections 5 and 6 provide some more background information about the two practices and describe some illustrative examples from the data, remarking on both similarities and differences in the polysemiosis of the two cultural practices. Finally, in Section 7 we summarize the contributions of the study and point to some directions for future research.

## 2 Our cognitive-semiotic framework

### 2.1 Signs, sign systems and polysemiotic communication

Definitions of terms like “sign” and “narrative” differ profoundly, and unfortunately there is much crosstalk both within and across disciplines like semiotics and narratology that employ these notions (Louhema et al. 2019; Ryan 2007; Sonesson 2007; Stampoulidis 2019). Here we provide our definitions in a rather postulative way, without arguing for their adequacy at length, but with references to works where this is done. More importantly, we intend to show the adequacy of these concepts through their productive operationalizations, provided in the following sections.

---

<sup>1</sup> Some boys engage as storytellers, but they are not encouraged to take up the practice in the same way as girls and more typically serve as onlookers.

A fundamental distinction is that between *signs* and *signals*. Both of these elements of communication consist of pairings of *expressions* and *meanings*, but only signs are used to denote things, properties, or events (i.e., intentional objects), with the producers or interpreters of the signs being at least to some degree aware of this relation, as explicated in the definition offered by Zlatev et al. (2020: 160): “A sign  $\langle E, O \rangle$  is used (produced or understood) by a subject S, if and only if: (a) S is made aware of an intentional object O by means of expression E, which can be perceived by the senses. (b) S is (or at least can be) aware of (a).”

Most (though not all) words, pictures, and gestures – as used by people, and possibly by some linguistically trained animals (Savage-Rumbaugh and Lewin 1994) thus qualify as signs. On the other hand, unintentional signals of emotion and other kinds of (social) meaning such as yawning, laughter, cries of pain, produced and interpreted with little if any conscious awareness (Burling 2005), may fulfill (a) but not (b) in the definition above, and do not qualify as signs.

A second fundamental concept is that of a *semiotic system*, which consists of (a) all the signs or signals of a particular type, and (b) their interrelations. Signals, like spontaneous facial expressions, also form systems, but these are much more constrained in their meaning potential and complexity than sign systems. They are *closed* as opposed to *open* systems, which can be constantly extended with new signs and sign combinations (Arbib 2005). Thus, the term “semiotic system” is a hypernym, while “sign system” and “signal system” are hyponyms, using lexical-semantic terminology (e.g., Saeed 2016). Focusing on sign systems we can ask: what delineates one sign system from another? The following six criteria help answer this question. Criteria (1–3) have to do with the *materiality* (i.e., the manner of producing and perceiving) of the signs, while (4–6) with their *semiotic potential* (i.e., the nature of how meaning-production is organized):

1. *Production*: the way the physical carrier of the sign, e.g., sound waves, marks on a surface, bodily movements, etc., is made.
2. *Modality*: the (predominant) sensory modality used to perceive the media in question, e.g., vision for pictures, hearing for speech and music, etc.
3. *Degree of permanence*: constraints on the duration of perception and interpreting of the signs in question, i.e., the reverse of what is sometimes called “fading,” with so-called *rapid fading* being a “design feature” of speech (Hockett 1960), but not (relatively speaking, even if the medium is as impermanent as sand) of writing.
4. *Double articulation* (or “duality of patterning”): some, but not all, signs can be constructed through systematic combinations of elements that are meaningless in themselves, e.g., phonemes in spoken languages (Jakobson 1965), or “cheremes” in (some, see below) signed languages (Tamura and Kawasaki 1988).

5. *Semiotic ground dominance*: the semiotic *ground* is the type of relation that exists between an expression and meaning (Sonesson 2010). Following the influential semiotic typology of Peirce, but applied to grounds rather than signs *per se*, three main kinds of ground can be defined: (a) *iconicity*, i.e., resemblance that can vary from imagistic (e.g., photographs) to abstract/relational (e.g., maps), (b) *indexicality*, i.e., spatio-temporal contiguity (closeness) or part-whole relations, and (c) *symbolicity*, i.e., social agreement (convention). A key insight attributed to Peirce but perhaps most clearly formulated by Jakobson (1965) is that the three kinds of ground are not mutually exclusive and typically co-exists in any sign. Still, it is possible (in most cases) to single out one of these grounds as being *predominant* for interpretation, e.g., convention in words, contiguity in natural symptoms, iconicity in representational pictures.
6. *Syntagmatic relations*: structuralist linguistics and semiotics characterized combinations of signs in larger units, such as a phrase, sentence, or text, in terms of “horizontal” (linear) combinations, complemented by paradigmatic (“vertical”) relations with other signs that could potentially take their place in this “slot” (Chandler 1994). In the case of language, such *structuralist* analyses are still productive (Sahlgren 2006), but it has been strongly argued that this cannot be generalized to other sign systems, in particular to those where iconicity is the dominant semiotic ground. Thus, forcing a structuralist analysis onto pictures amounts to a form of “linguistic imperialism” (Sonesson 1989) that distorts the phenomenon analyzed. Acknowledging this, we can still use the notion of syntagmatic relation as a criterion for discerning different types of sign systems, but need to regard this as a variable, distinguishing relations of more and less *compositional* kinds: where the meanings of the component signs combine in systematic, if not deterministic, manner.

With the help of these criteria, we can distinguish three fundamental and universal (in the case of depiction, at least in the case of picture comprehension) human sign systems: *language*, *gesture*, and *depiction*, and in addition: three subsystems for language, as shown in Table 1.

Starting with the sign system of *language*, criteria (1), (2), and (3) allow us to clearly distinguish its three sub-systems: speech, writing, and signed language. Writing has a relatively high degree of permanence – even if it is realized in “impermanent” media like paper, white board, or even sand – since the produced signs are made available to prolonged or repeated perception for at least some duration of time. Signed language is identical in terms of production and modality with gesture: produced by the whole body (and not just by the hands, as often erroneously assumed) and perceived predominantly through vision. Irrespective of these differences, the semiotic features of these sub-systems, i.e., criteria (4), (5), and

**Table 1:** Defining language, gesture and depiction as distinct sign systems, differentiated by properties of materiality (production, modality, permanence) and semiotic potential (double articulation, semiotic ground dominance, syntagmatic relations).

Criteria	Sign systems				
	Language			Gesture	Depiction
	Speech	Writing	Signed language		
(1) <i>Production</i>	Vocal	Extra-bodily	Whole body	Whole body	Extra-bodily
(2) <i>Modality</i>	Auditory (+Visual)	Visual (+Tactile)	Visual (+Tactile)	Visual (+Tactile, Auditory)	Visual (+Tactile)
(3) <i>Permanence</i>	Low	High	Intermediate	Intermediate	High
(4) <i>Double articulation</i>	Yes		Yes/No	No	No
(5) <i>Semiotic ground</i>	Symbolic > Iconic, Indexical			Iconic, Indexical > Symbolic	Iconic > Indexical, Symbolic
(6) <i>Syntagmatic relations</i>	Compositional (systematic)			Linear	Possibly linear

(6) are essentially the same. All three have *double articulation*: phonemes or graphemes combine systematically to form meaningful morphemes, and this is also the case for some but not all signed languages (Sandler 2012). All human languages are characterized by high degrees of conventionality (Clark 1996) and normativity (Itkonen 2003), making symbolicity their predominant semiotic ground, even if iconicity (Devyllder 2018; Dingemanse 2012), and indexicality (Kravchenko 2017) are also essential. This means that the signs of language are not “arbitrary” (Ahlner and Zlatev 2010; Jakobson 1965; Zlatev 2014), despite common claims to the contrary, in the wake of the ambiguous use of the term by one of the canonical texts of linguistics and semiotics, Saussure’s posthumously compiled and published *Cours de linguistique générale* (Saussure 1916). The *syntagmatic relations* (criterion 6) between the signs of language are characterized by a high degree of compositionality, where the meaning of a composite sign is built up (at least in part), from the meanings of its constituent signs, and the rules for combining these – though not in a mechanical “building block” manner, given the context-sensitivity of linguistic meaning (Zlatev 1997).

The everyday term “gesture” is highly ambiguous – as basically all semiotic notions – and could be used to cover all from postures and involuntary movements in so-called “body language,” to the conventionalized signs of signed languages. This is, of course, highly confusing. In our cognitive-semiotic framework, the signs of the

sign system of *gesture* are essentially understood in the sense of Kendon (2004) as bodily “movements that partake of . . . features of manifest deliberate expressiveness to an obvious degree” (2004: 14), complemented by the definitions offered by Andrén (2010), who pointed out that (as a sign system) gesture should be distinguished both from signed languages (“the upper limit”) and postures or practical actions (“the lower limit”). Silently miming, sometimes called “pantomime,” is thus a form of gesture (Zlatev et al. 2020). It is precisely a high degree of conventionality/normativity that distinguishes signed language from gesture, even though there is a large class of gestures, often called “emblems” such as the OK or VICTORY signs, where symbolicity is indeed the dominant semiotic ground (Zlatev and Andrén 2009). For most gestures, however, it is their resemblance (iconicity) and attention-orienting (indexicality) functions that determine how they are produced and interpreted (Zlatev 2015b), and this may even apply to some conventionalized gestures with roots in *bodily mimesis* (Müller 2016).

The third sign system to be differentiated is that of *depiction*, by which we mean a system of signs where each is a shape inscribed upon a (largely) two-dimensional surface that represents a (three-dimensional) object, an action, a whole event, or even more abstract notions such as shapes and relations.<sup>2</sup> While conventions on how to perform depiction signs have varied immensely, from the caves of Lascaux to the paintings of Picasso, the dominant semiotic ground in depiction (though notably not in *art* in general) is that of iconicity, including of the more abstract, diagrammatic type (Sonesson 1989). Analogously to the case of the dominant symbolicity of language, one needs simultaneously to acknowledge the necessary presence of various forms and degrees of indexicality (e.g., the setting in which a picture is viewed) and symbolicity (e.g., the conventions of how certain types of pictures, like icons, are produced).<sup>3</sup>

The signs of the systems of gesture and depiction can be analyzed into smaller sign *units* and *nuclei* (Green 2014; Kendon 2004), as we show in Section 3. These, however, are not made up of minimal distinctive elements like phonemes or graphemes, and hence lack double articulation. Further, these sign systems have much less systematic manners in arranging sequences of signs, making it more difficult (but not impossible) to express complex messages such as *narratives* (Donald 1991; Ryan 2012). Yet gesture, for example when realized as (the core of) pantomime, can be used to narrate simple chronological stories, due to the

---

2 Note that this is distinct from the use of the term “depiction” in gestures (Clark 1996) or in ideophones (Dingemans 2012).

3 Orthodox Christian *icons* are thus poor examples of iconic signs, just as many visual “symbols” are only somewhat symbolic, in the technical sense in which these semiotic terms are here used, in the footsteps of Peirce, without attempting a full overlap with his specific (and changing) concepts.

temporal sequences of the gestural signs (Zlatev et al. in press). This is not the case for single pictures, which can only narrate in a *secondary* manner, once the original stories are already known (Stampoulidis 2019). And of course, it is necessary to form a convention of how to “read” a sequence of pictures for them to narrate in a primary way (hence, “possibly linear” in the corresponding cell in Table 1).

The two sign systems of gesture and depiction are also clearly distinct in terms of their materiality. While vision is in both cases the dominant modality, the fact that gesture is produced by the whole body and does not leave a trace implies differences in permanence. In general, gesture has a degree of permanence that is intermediate between speech, on the one hand, and depiction and writing, on the other. This is due to the mid-air *hold* elements of many gestural expressions, used when the need for emphasis or audience attention requires it (see Section 3.3). There are no inherent differences in permanence between the systems of depiction and writing, with specific differences having to do with the medium in which pictures/texts are produced (e.g., stone vs. paper vs. sand). Written texts, like the Rosetta Stone, have been for historical reasons more often “made to last” than pictures, though of course there are many examples where the contrary is the case.

Having defined the three universal sign systems in terms of the set of criteria, we need to emphasize two things. The first has already been mentioned (see footnote 2): to define language, gesture and depiction in this way requires a considerable degree of abstraction and generalization. In actual life, each system is realized in different *cultural practices*, with their specific properties. For example, depiction can be realized in practices that are as distinct as oil painting, photography, and sand drawing. Gesture can be realized in charades, formalized pantomime, spontaneous gesturing while speaking etc. And the “language games” in which language may be realized are limitless (Wittgenstein 1953). We claim, however, that the systems can be defined and distinguished with the help of essential features on the *eidetic level* (Husserl 1989), corresponding to some degree to the *universal level* (as opposed to the historical and situated levels) of Integral Linguistics (Cosieriu 1985, 2000). Of course, we need to delve into specific practices and media (like film) when we need to analyze in detail how these systems are realized and interact with one another.

This brings us naturally to the second major theoretical point: human communication is very seldom *monosemiotic*, i.e., based on a single sign system, for example in practices that involve (pictureless) books, uncommented paintings, or completely silent pantomime. In fact, unless some particular genre-based constraints are imposed, human communication is as a rule *polysemiotic*: it combines fluently different semiotic systems, both sign systems and signal systems. For example, human face-to-face communication, the most ancient and still prototypical form of human communication (Clark 1996), typically combines the sign



system of language (realized as speech or signed language), and the sign system of gesture, along with the signal systems such as face expressions and postures. And as we know, for example from the two cultural practices under study, this is also very often supplemented by depiction. Thus, the abstraction of sign systems that our framework offers is made for analytic purposes: for the sake of reaching a better appreciation of the composite, we need a better understanding of the parts.

## 2.2 Alternative frameworks

### 2.2.1 The multiple senses of “multimodality”

The currently popular term “multimodality” has sprung forth in different fields, from cognitive psychology to conversation analysis, and hence the senses in which scholars in each field use it differ substantially. Even in a single field like Cognitive Linguistics (CL), it is used in different and conflicting senses, as pointed out by Devylder, in a review of a recent handbook:

Defining something as multimodal naturally implies that it involves at least two modalities, or modes ... For Vandelanotte (p. 158), a modality is a sensory channel (e.g. “the visual modality”), for Sullivan (pp. 389–391) in line with the CL tradition that uses the term ‘visual metaphors’ in contrast to ‘linguistic metaphors,’ modality has to mean that pictures and written text are two modalities (the combination of which is thus “multimodal”), for Feyaerts et al. (p. 135), the term extends to just about any aspect of a face-to face interaction ... [Thus] when one reads a CL paper on multimodality, one can expect the term to either mean: the combination of text and image, the combination of gesture and speech, the combination of vision and hearing, or a combination of all of the above and more. (Devylder 2019b: 149–150)

A main problem is thus that the term “modality” is used either in the sense of (a) perceptual faculty, or (b) in the sense of semiotic system, without distinguishing sign from signal systems, and (c) often even more broadly, to include what we called above cultural practices:

Metaphors are expressed in different modalities, ranging from words and gestures, to music, pictorial advertisements, pieces of art, including films and audiovisual compositions more generally. (Müller 2017: 300)

Such conflation of perceptual and semiotic dimensions is unfortunate. Some, who recognize this, distinguish between “multimodality in the strict sense” (Vandelanotte 2017: 161) for different sensory (and productive) channels, and “in the loose sense” for different “semiotic resources” like text and image. But it is unclear why we should use the term for both of these very different phenomena. Our cognitive-semiotic framework allows clearly distinguishing perceptual modalities and their

combinations (i.e., *multimodality*) and semiotic systems and their combinations (i.e., *polysemiosis*). Indeed, Table 1 shows that there is no direct correspondence between semiotic system and perceptual modality. For example, speech is typically perceived not only as sound, but visually as mouth movements, as shown by the well-known McGurk effect (McGurk and MacDonald 1976). Various manifestations of depiction such as reliefs can be perceived through multiple modalities. Thus, it is simply confusing to speak of “visual communication” without additional clarification, as vision is the dominant sensory modality for all three sign systems – in reading written text and perceiving pictures or gestures – while it plays a secondary role in the sub-system of speech (see Table 1).

Another way to attempt to salvage the notion of multimodality is to state that what it combines are not “modalities,” but “modes.” But this notion has been even harder to define in a consistent way, and it has in fact often been left undefined. Acknowledging this but faced with the need to analyze different aspects of “multimodal metaphors” in Forceville (2017: 27) states: “I distinguish the following modes: written language, spoken language, visuals, music, non-verbal sound, gestures, olfaction, taste, and touch.” Once again, different semiotic systems like language, gesture and music, sub-systems of language (speech and writing), and different sensory modalities are conflated. Taken to an extreme, a “mode” can be just about anything that the analyst deems meaningful in social interaction, leading to the conclusion that “there is no theoretical limit to the number of modes that may be recognized in various socio-cultural contexts, and this leads to an abundance of modes that are difficult to compare” (Green 2014: 9–10). Such considerations call for a more constrained approach.

### 2.2.2 Green’s V-units

The work of Green (2014) has been strongly inspirational for our own approach. Lead by a similar critique of the notion of “multimodality” as expressed above, Green arrives at the conclusion that two modalities/channels are (for the most part) sufficient for analyzing the Central Australian cultural practice of sand stories: the visual and the auditory. While acknowledging that gesture and depiction constitute different semiotic systems, Green (2014: 75) maintains that: “many of the interesting aspects of these stories are in fact more ‘hybrid’ in nature and occur between and across the boundaries of one semiotic system and another.” For example, the transition between a drawing and a gesture often occurs smoothly as the hand changes its position from the ground to the air. Another example is so called “drawing beats,” which “are like beats [i.e., a type of gesture] in the manner of their articulation, but they also leave marks on the ground that have an iconic or representational function” (2014: 219). Motivated by such considerations and by

extending Kendon's (2004) notion of *gesture unit* (G-unit), Green defines a *visual unit* (V-unit), which includes all instances of communicative visible bodily-action, irrespective of whether it is performed on the ground or in the air (Green 2014: 79).

While we agree that there is a high degree of interaction between the sign systems of gesture and depiction in practices such as sand drawing, we would argue that is both possible and (at least in some cases) beneficial to distinguish between them in analysis. As we show in the following sections, the various units (signs) of depiction and gesture seldom fully overlap in time and are often complementary in terms of semiotic grounds. One type of behavior that Green (2014) viewed as combining gesture and depiction is the “drawing beats” commonly seen in Central Australian sand stories. In Section 3.4, we discuss how we approached analyzing these as only gesture, only depiction, or both.

## 2.3 Summary

In this section we presented our general cognitive-semiotic framework, defining the key concepts of signs and signals, semiotic systems (falling into sign systems and signal systems) and polysemiotic communication (polysemiosis), briefly contrasting it with other frameworks that do not clearly distinguish between polysemiosis and the interactions of different sensory channels, which is how we define multimodality in our approach. But given that “the proof of the pudding is in eating it,” it is now time to turn to our coding system of polysemiotic communication and its application to the two sand drawing practices under investigation.

## 3 A coding system for analyzing polysemiosis

The coding system was developed with the help of the conceptual-empirical loop of cognitive semiotics (Stampoulidis 2021; Zlatev 2015a). This implies that the theoretical framework presented in Section 2 was used to guide general decisions, such as distinguishing between the “tiers” of gesture and depiction, and the three kinds of semiotic grounds (iconic, indexical, symbolic). On the other hand, the precise decisions on how to *operationalize* these were made only after extensive work with the material, and numerous discussions among the investigators. The material involved systematic analysis of 23 video-recordings of sand drawing performances on Paama, Vanuatu and 20 sand stories in Central Australia, two cultural practices described in detail in the following sections. The coding system was implemented in the ELAN video annotation software (Wittenburg et al. 2006). We illustrate some of the key notions with the help of ELAN screenshots, though without including

unnecessary technical details, such as the precise structure, and the dependencies between the tiers.

The 43 video files that we collected varied from 1:02 to 23:54 min of length, with average 05:49. Each showed one or more participants of the respective Indigenous culture who performed a sand story narration (in Central Australia) or drew and commented on a sand drawing (on Paama), on the request of a Western observer. Each polysemiotic *performance* was video recorded with two cameras, one from above and one from the side, as employed by Green (2014).

Each performance was analyzed for the three main sign systems (implemented as “main tiers”): speech, depiction, and gesture (see Section 2). The expressions on each system were analyzed in terms of several levels of structural complexity, corresponding to (approximately) the levels of (a) whole *episodes*, understood as a sequence of a limited number of (causally) interrelated events, (b) *events*, and (c) *objects, actions, or properties*. As shown in Table 2, not all of these levels were used for all semiotic systems, and the notion of G-unit, defined mostly based on expression criteria (see below), was allowed to vary in complexity between whole episodes and individual events.

Given that signs are pairings of expressions and meanings (see Section 2), signs on all levels were defined on the basis of both *expression* (sometimes called “formal”) criteria and *semantic* criteria, even if in some cases expression criteria were more important (e.g., for defining G-units) while in other cases, semantic ones were the dominant ones (e.g., for defining morphemes). In the following three sub-sections, we explain the way the signs of each semiotic system were segmented and analyzed, before returning to the system-blending phenomena of “wire-tapping” and “drawing beats” (Section 2.2.2) in Section 3.4.

### 3.1 Language (speech)

The stream of speech of each interlocutor was segmented in *utterances*, motivated by the fact that each utterance constitutes the “minimal move” in a conversation or

**Table 2:** Levels of analysis for the three semiotic systems.

<b>Sign system</b>	<b>Language (speech)</b>	<b>Depiction</b>	<b>Gesture</b>
<i>Level of complexity</i>			
<i>Episode</i>		D-frame	G-unit
<i>Event</i>	Utterance	D-phrase	
<i>Object/action/property</i>	Morpheme	D-nucleus	G-nucleus

language game (Zlatev 1997). The expression criteria for segmentation were intonation and pauses, along with the semantic criterion that the utterance needed to be interpretable as denoting a particular event, thing, or property. For example, Figure 1 shows three such utterances.

Each utterance was further segmented into *morphemes*: with either lexical or grammatical meanings. Naturally, we could have opted for intermediary (i.e., phrase) levels of analysis, but since the system of language was not the primary focus of the study, these two structural levels were sufficient for our purposes. English translations were given to both the whole utterances and to each morpheme, in the latter case a *gloss*. The predominant semiotic ground of each morpheme was also coded. This was considered to be symbolic (i.e., conventional) by default (see Section 2.1 and Table 1), but in case the morpheme/word was an *ideophone* (Dingemanse 2012) the ground was marked as iconic, and if it was a *demonstrative expression* (Diessel 2006) it was coded as indexical.

## 3.2 Depiction

Since the cultural practices analyzed were chosen precisely because they involve depictions, we paid considerable importance to this semiotic system. As shown in Table 2, three levels of structural complexity were operationally distinguished, one level more than for the other two systems.

Grid	Text	Subtitles	
1	00:00:33	00:00:41	00:00:06
2	00:00:46	00:00:53	00:00:07
3	00:01:00	00:01:07	00:00:07
4	00:01:04	00:01:07	00:00:03
5	00:01:09	00:01:11	00:00:02
6	00:01:12	00:01:13	00:00:01
7	00:01:26	00:01:28	00:00:02
8	00:01:44	00:01:46	00:00:02
9	00:01:56	00:01:57	00:00:01
10	00:01:57	00:01:58	00:00:01
11	00:01:58	00:01:59	00:00:01
12	00:02:02	00:02:03	00:00:01
13	00:02:03	00:02:07	00:00:04
14	00:02:07	00:02:08	00:00:01
15	00:02:12	00:02:14	00:00:02
16	00:02:24	00:02:27	00:00:03
17	00:02:29	00:02:30	00:00:01
18	00:02:30	00:02:31	00:00:01
19	00:02:34	00:02:37	00:00:03
20	00:02:38	00:02:41	00:00:03
21	00:02:44	00:02:46	00:00:02
22	00:02:48	00:02:49	00:00:01
23	00:02:54	00:02:55	00:00:01
24	00:02:56	00:02:57	00:00:01

TIME	TEXT	GLOSS	SEMANTIC	GROUND
00:01:57:000	me	me	me	me
00:01:58:000	me	me	me	me
00:01:58:500	me	me	me	me
00:01:59:000	me	me	me	me
00:01:59:500	me	me	me	me
00:02:00:000	me	me	me	me
00:02:00:500	me	me	me	me
00:02:01:000	me	me	me	me
00:02:01:500	me	me	me	me
00:02:02:000	me	me	me	me
00:02:02:500	me	me	me	me
00:02:03:000	me	me	me	me

Figure 1: Utterances and morphemes within the sign system of language (speech).

### 3.2.1 Depiction frames (D-frames) and depiction phrases (D-phrases)

Using expression criteria, the D-frame was individuated by the performer making two *wipes*: striking out with the hand or stick everything that is drawn on the sand. More specifically, each D-frame started with the first D-phrase (either *produced* or *commented*, see below) and ended with the final D-phrase before a wipe.

Semantically, the D-frame corresponds to an *episode* (which could be considerably long), consisting of one or more events. The role of semantic criteria was essential in the case of “partial wipes”: when the speaker wiped some, but not all the markings. Typically, this was not taken to end the D-frame but only the D-phrase they occur in, as partial wipes appeared to have the function of refining the topic, rather than changing it.<sup>4</sup> Interestingly, it was typical for the Paamese sand drawings to consist of only one D-frame, while in the Pitjantjatjara sand stories, there were usually several D-frames, of lengths varying from a couple of seconds to several minutes (see Section 5).

Depiction phrases (D-phrases) were typically (but not always) of smaller length and complexity than D-frames. Three kinds of D-phrases were differentiated and coded: (a) *produced*: where the image in the sand was drawn co-temporally (see Figure 2), (b) *commented*: where the image was drawn earlier, and was now being commented (see Figure 3), (c) *mixed*: a combination of (a) and (b), see Figure 4

The screenshot shows a video player interface. The main video area displays a scene of several people sitting on the sand in an outdoor setting, engaged in drawing. On the right side, there is a subtitle list with columns for 'Grid', 'Text', 'Subtitles', and 'Lexicon'. Below the video player, a detailed subtitle table is visible, showing the structure of the subtitle tracks.

Track	Start Time	End Time	Text
1	00:00:06.000	00:00:08.000	DepictionFrame@DEPICTION_FRAME@SPEAKER1
2	00:00:08.000	00:00:09.000	Comment@DEPICTION_FRAME@SPEAKER1
3	00:00:09.000	00:00:10.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
4	00:00:10.000	00:00:11.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
5	00:00:11.000	00:00:12.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
6	00:00:12.000	00:00:13.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
7	00:00:13.000	00:00:14.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
8	00:00:14.000	00:00:15.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
9	00:00:15.000	00:00:16.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
10	00:00:16.000	00:00:17.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
11	00:00:17.000	00:00:18.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
12	00:00:18.000	00:00:19.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
13	00:00:19.000	00:00:20.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
14	00:00:20.000	00:00:21.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
15	00:00:21.000	00:00:22.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
16	00:00:22.000	00:00:23.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
17	00:00:23.000	00:00:24.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
18	00:00:24.000	00:00:25.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
19	00:00:25.000	00:00:26.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
20	00:00:26.000	00:00:27.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
21	00:00:27.000	00:00:28.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
22	00:00:28.000	00:00:29.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
23	00:00:29.000	00:00:30.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
24	00:00:30.000	00:00:31.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
25	00:00:31.000	00:00:32.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
26	00:00:32.000	00:00:33.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
27	00:00:33.000	00:00:34.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
28	00:00:34.000	00:00:35.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
29	00:00:35.000	00:00:36.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
30	00:00:36.000	00:00:37.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
31	00:00:37.000	00:00:38.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
32	00:00:38.000	00:00:39.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
33	00:00:39.000	00:00:40.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
34	00:00:40.000	00:00:41.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
35	00:00:41.000	00:00:42.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
36	00:00:42.000	00:00:43.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
37	00:00:43.000	00:00:44.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
38	00:00:44.000	00:00:45.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
39	00:00:45.000	00:00:46.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
40	00:00:46.000	00:00:47.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
41	00:00:47.000	00:00:48.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
42	00:00:48.000	00:00:49.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
43	00:00:49.000	00:00:50.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
44	00:00:50.000	00:00:51.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
45	00:00:51.000	00:00:52.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
46	00:00:52.000	00:00:53.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
47	00:00:53.000	00:00:54.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
48	00:00:54.000	00:00:55.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
49	00:00:55.000	00:00:56.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
50	00:00:56.000	00:00:57.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
51	00:00:57.000	00:00:58.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
52	00:00:58.000	00:00:59.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
53	00:00:59.000	00:01:00.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
54	00:01:00.000	00:01:01.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
55	00:01:01.000	00:01:02.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
56	00:01:02.000	00:01:03.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
57	00:01:03.000	00:01:04.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
58	00:01:04.000	00:01:05.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
59	00:01:05.000	00:01:06.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
60	00:01:06.000	00:01:07.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
61	00:01:07.000	00:01:08.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
62	00:01:08.000	00:01:09.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
63	00:01:09.000	00:01:10.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
64	00:01:10.000	00:01:11.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
65	00:01:11.000	00:01:12.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
66	00:01:12.000	00:01:13.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
67	00:01:13.000	00:01:14.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
68	00:01:14.000	00:01:15.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
69	00:01:15.000	00:01:16.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
70	00:01:16.000	00:01:17.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
71	00:01:17.000	00:01:18.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
72	00:01:18.000	00:01:19.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
73	00:01:19.000	00:01:20.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
74	00:01:20.000	00:01:21.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
75	00:01:21.000	00:01:22.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
76	00:01:22.000	00:01:23.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
77	00:01:23.000	00:01:24.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
78	00:01:24.000	00:01:25.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
79	00:01:25.000	00:01:26.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
80	00:01:26.000	00:01:27.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
81	00:01:27.000	00:01:28.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
82	00:01:28.000	00:01:29.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
83	00:01:29.000	00:01:30.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
84	00:01:30.000	00:01:31.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
85	00:01:31.000	00:01:32.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
86	00:01:32.000	00:01:33.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
87	00:01:33.000	00:01:34.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
88	00:01:34.000	00:01:35.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
89	00:01:35.000	00:01:36.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
90	00:01:36.000	00:01:37.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
91	00:01:37.000	00:01:38.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
92	00:01:38.000	00:01:39.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
93	00:01:39.000	00:01:40.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
94	00:01:40.000	00:01:41.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
95	00:01:41.000	00:01:42.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
96	00:01:42.000	00:01:43.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
97	00:01:43.000	00:01:44.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
98	00:01:44.000	00:01:45.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
99	00:01:45.000	00:01:46.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1
100	00:01:46.000	00:01:47.000	DepictionPhrase@DEPICTION_FRAME@SPEAKER1

Figure 2: A produced D-phrase in the sign system of depiction.

4 However, if the speaker made what looked like a partial wipe, with some markings left, but not referenced again and deleted as part of a later wipe, then such a wipe was considered a full wipe, and thus the end of the D-frame.





wipe, marked its end. The semantic criterion was that it must depict a meaningful whole that could be identified (and labelled) based on one or more of the three kinds of evidence:

- a) *Primary iconicity* (Sonesson 1997): an image that shows the object that it depicts so transparently that it is possible to recognize it without further evidence.
- b) *Co-depiction speech* (and possibly also co-depiction gesture): the depicted object is being named by the performer or some other participant.
- c) *Culture-specific convention*: A sedimented representation, documented from previous cases of (b), or from general ethnographic knowledge.

Given that iconicity is the default semiotic ground in depiction (see Section 2), coding (a–c) was informative not so much about the presence or absence of iconicity, but about the degree and type of it. The D-phrase was given a label in English (e.g., “soakage” and “co-depiction speech” in Figure 2) and the type of evidence was provided.

### 3.2.1.2 Commented D-phrase

The same semantic criterion as for produced D-phrases was used, but naturally, the expression criterion had to be different: it was required that the speaker produce one or more gestures and/or speech expressions which called attention to the depiction-based sign. Further, we requested that these must all occur within the same (a) utterance or (b) G-unit (see below). Each commented D-phrase was coded in the same manner as for produced ones, as described above.

### 3.2.1.3 Mixed D-phrase

This was the case where there were one or more produced D-nuclei (see below) combined with one or more commented D-nuclei, and these were also parts of the same meaningful whole. For example, in the *manun* (‘flying fox’) sand drawing performance shown in Figure 4, the sand drawer describes how the animal hangs and swings on the branch of breadfruit trees to eat (i.e., “the meaningful whole” event). To augment his verbal description, the sand drawer first points to the already drawn foot of the animal (that he calls *hook*): a *commented* D-nucleus in our coding system. Secondly, but still within the same meaningful whole event (i.e., D-phrase), he draws an additional stroke in the ground that denotes the branch that the flying fox grip to swing: a *produced* D-nucleus.

## 3.2.2 Depiction nuclei (D-nuclei)

The smallest parts of depictions in the analysis were the D-nuclei, defined as meaningful parts of the whole represented/depicted by the D-phrase they are a part of. While it was possible for D-phrases to consist of a single D-nucleus (as in Figures 2



and 3), it was common for the performers to explicitly distinguish such parts (e.g., “the hook,” “the breadfruit tree” in Figure 4) within a single D-phrase (irrespective of type: produced, commented, mixed).

The expressive and semantic criteria used for defining D-nuclei were like those for D-phrases. For produced D-nuclei this was “continuous spatial marking”: it must be possible to draw the nucleus with one continuous line, without tracing back along the existing lines.<sup>5</sup> The semantic criterion was that it must depict an identifiable object (a part if there are several D-nuclei) based on one or more of the same criteria as for D-phrases: (a) primary iconicity, (b) co-depiction speech, (c) documented culture specific convention. For a commented D-nucleus, the expressive criterion was that the performer produced an expression in the semi-otic systems speech and/or gesture that brings attention to the depicted object/part.

Thus, the type of evidence for interpreting the depictive nature of the sign was used twice whenever a D-phrase consisted of multiple D-nuclei, and thus of a represented whole and its parts.<sup>6</sup>

### 3.3 Gesture

The semiotic system of gesture has been previously analyzed into components extensively in the literature, usually based on the seminal work of Kendon (2004). We assumed the same point of departure, but used a somewhat simplified system, distinguishing only between gesture units (G-units) and gesture nuclei (G-nuclei), as shown in Table 2.

#### 3.3.1 Gesture units (G-units)

Gesture units were defined on the basis of the following expression criteria:

- a) Each G-unit should start from a place of rest (e.g., hands in lap) and finish at a place of rest.
- b) It begins with the *preparation phase* of the first G-nucleus (see below): when the articulator starts to move away from place of rest.

---

<sup>5</sup> The nucleus did not need, however, to have been produced in a single uninterrupted line – the drawer may have paused or lifted the drawing tool and continued from the same, or a different place.

<sup>6</sup> For example, in Paamese *Hoktsenfai* (‘shark spirit’) drawings, a wavy line drawn was described by the speaker as being the shark spirit’s “propeller.” Thus, the label for this D-nucleus was “propeller,” and the evidence: (b) co-depiction speech. There was a total of five such signs in the performance. In later occurrences, the evidence for the label “propeller” assigned to these signs was (b) co-depiction speech, if they were named again, or (c) convention, if they were not.

- c) It ends at the end of the *retraction* of the last G-nucleus within the unit. Even a very brief return to a place of rest marks the end of a G-unit.

A G-unit does not necessarily start and end at the same place of rest and may also depart from or end in some other kind of expression, such as drawing, or a self-regulator (e.g., scratching oneself). In the former case, the G-unit begins when the speaker lifts their finger/tool from the drawing space and ends when the finger/tool back is brought back to the surface.

### 3.3.2 Gesture nuclei (G-nuclei)

Each G-unit must contain one or more G-nuclei, the semantic core(s) of the G-unit, thus confirming the principle that sign identification requires both expression and semantic criteria. With respect to expression, a G-nucleus could be (a) a *stroke*, possibly together with a post-stroke hold: a moment of rest in the air typically occurring after the stroke, or (b) a *simple hold*. These components were operationally defined as follows:

- The *stroke* starts at the first frame of the motion. It can start from either the end of a preparation movement, a pre-stroke hold, directly from the end of a previous stroke, or a place of rest. It ends at the last frame of the motion, or, if there was a post-stroke hold, at the last frame of the hold and can be followed either by a retraction or a preparation for a new stroke or end at the place of rest.<sup>7</sup>
- The *simple hold* starts at the first frame where the movement reaches its apex (i.e., the point of furthest reach) and ends at the last frame at the apex, before the speaker retracts the articulator from that location.

In cases of repeated movements, it was challenging to distinguish between single and separate strokes. It is generally the case that the size of repeated movement may gradually change, so change in movement size alone cannot be evidence for coding repeated movements as separate strokes. When determining if a repeated movement should be coded as one single stroke (G-unit), or multiple ones, the following principles were applied:

---

<sup>7</sup> When deciding which part of a bi-directional movement should be coded as stroke, and which part is preparation and/or retraction, the rule was that the downwards part of the movement shall be marked as the stroke, unless the upwards part of the movement is physically marked (through speed, hand shape, etc.) or clearly aligned with a stressed syllable while the downwards part is not. To decide how much of an arch movement should be included as stroke, the rule was that the whole arch should be coded as stroke, if it is performed as one smooth movement.

- Singular strokes: repeated movements consisting of a fluid motion with no holds at direction-shifts, and repeated movements following along one single axis, e.g., *zigzags* or *spirals*.
- Separate strokes: Repeated movements that were either rhythmically aligned with syllables of the co-gesture speech or aligned with separate words.

Cases of moving the hand with an index-finger configuration between index-finger pointing gestures were not considered as a meaningful part of a G-nucleus. If on the other hand, the movement between two points was repeated, then the movement was considered as a meaningful part of the stroke.

Given that each G-nucleus was a meaningful sign, it was coded with a gloss and dominant semiotic ground. This could be iconic, as in Figure 5, indexical, including acts calling attention to D-phrases (making the latter a commented D-phrase) as in Figure 6, or symbolic (conventional).

As with the depiction system, we marked the type of evidence used for deciding on the meaning of iconic gestures: (a) primary iconicity, (b) co-gesture speech, (c) culture specific convention. For example, Figure 5 shows a Paamese sand drawer who explains verbally how a shark spirit, which can be summoned by the drawing of the geometrical pattern, would hold a victim with its fins. The drawer augments his verbal explanation with a G-unit with iconic ground, with evidence of the type primary iconicity, representing a taking & holding action. In contrast, Figure 6 shows a G-unit with predominantly indexical ground, interacting with a commented D-phrase.



**Figure 5:** An iconic G-unit, with evidence based on primary iconicity (a re-enacting gesture of taking and holding).



**Figure 6:** An indexical G-unit, drawing attention to a part of the drawing, a commented D-phrase.

### 3.4 Tapping: gesture, depiction or both

As pointed out in Section 2, one of the arguments for Green (2014) to conflate the sign systems of gesture and depiction into the notion of “visual units” were phenomena like “tapping” and “drawing beats,” which were hard to categorize as only gesture or as only depiction. For example, Green notes that “a story-wire may be used to prod or stab the earth to add rhetorical impact to a narrative, or it may be used to perform non-imagistic discourse functions – akin to beats” (2014: 75). At the same time, these may affect the surface of drawing space, and result in depiction. We addressed this difficulty by coding ambiguous actions like tapping based on the following criteria:

- *Only depiction*: if it resulted in a representation of distinct entities (e.g. family members), or movement trajectory (e.g., “going somewhere”).
- *Only gesture*: if the taps functioned as a marker of the genre of sand-drawing; or else like so-called “beat gestures,” used for emphasis and subtly drawing attention to parts of the speech, or for holding the floor and maintaining an active turn while preparing the next spoken utterance. In such cases, any traces on the sand were accidental.
- *Both depiction and gesture*: if both the visual mark on the drawing and the movements themselves served together to create the same function (e.g., the marks leading attention to an area of the drawing, and thus an indexical sign, like an “X” on a treasure map), but with the movements forming an essential part of the process of guiding the attention to the desired spot on the drawing.

When coded as gesture, tapping was part of a G-unit and functioned as a G-nucleus, like a stroke. When repeated, a single G-unit was coded from the first stage of

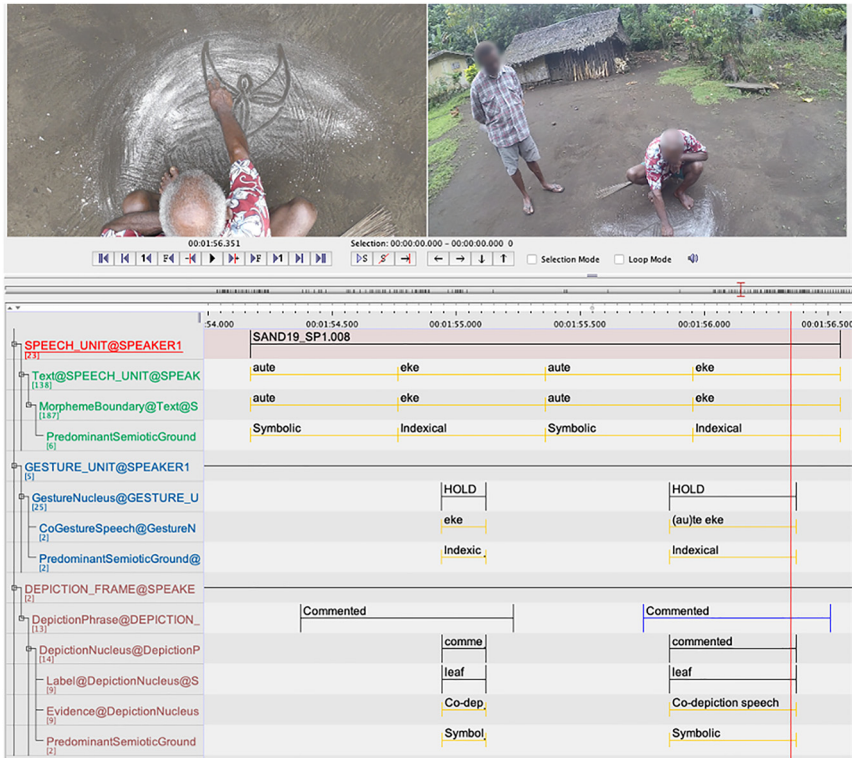
downwards movement towards the first tap to its end at the touch of the ground of the last tap. The gloss for the repeated tap was given as “tap” and the semiotic ground of such G-nuclei was coded as indexical if drawing attention, or as conventional, if (at least in part) marking the genre of sand-drawing. On the other hand, if tapping was performed on markings signifying locations, people, or objects, and co-occurred with hesitant or repeated speech, then each tap was coded as an individual stroke. In practice, the decision was based on timing: one smooth, repeated movement between two outer points was considered one stroke. If there were holds at the points, it was coded as several strokes.

## 4 Patterns of polysemiotic interaction and operationalizations

The previous section introduced the coding criteria from the three sign systems independently, and did so in quite some detail, as these criteria are meant to be explicit enough to provide for a high degree of reliability and replicability. But of course, the whole purpose of the analysis was to combine the different semiotic systems, and to establish patterns in their interaction. In the present section, we illustrate with an example from the Paamese data and spell out some operational decisions for the coding of depictions and gestures with some degree of iconicity. Figure 7 illustrates a segment of a Paamese sand drawing performance, with subtle interaction between the semiotic systems.

On the highest tier, SPEECH, the utterance is segmented into morphemes that are glossed and coded with semiotic grounds. The Paamese word *aute* signifies a leaf based on a linguistic convention sedimented in the Paamese lexicon. Accordingly, the predominant semiotic ground for this speech segment is coded as Symbolic. *Eke* (‘here’) is a locative noun that is used to draw the hearer’s attention to something present in the communicative context. The predominant ground for the speech segment was therefore coded as Indexical. In the timeframe during which the speech unit *aute eke, aute eke* ‘a leaf here, a leaf here’ is produced, two G-nuclei are also produced. They qualify as simple holds (HOLD) according to the criteria described in Section 3.3.2. Significantly, the gestures align in time with the two instances of the locative noun *eke*, and their predominant semiotic ground was likewise coded as Indexical, given that they draw the hearer’s attention to a specific part of the drawing.

Further, the Paamese sand drawer is pointing at a precise part of the sand drawing. Even though the depiction segment is not being produced during this temporal segment, it is an instrumental part of the communicative situation: the



**Figure 7:** Coding interaction between the three systems of speech, gesture, and depiction.

utterance ‘this is a leaf here’ with an accompanying pointing gesture would not make sense without the presence of the drawing. The depiction system therefore clearly plays a key role here, and accordingly the two D-phrases were coded as ‘Commented’ rather than ‘Produced’. Each D-phrase has its nucleus, glossed as ‘leaf,’ and of the type of evidence used to determine the depictive nature of the sign. To remind, these could be of three types: (a) primary iconicity, (b) co-depiction speech, (c) culture-specific convention (see 3.2.2). Here, the two parts of the drawing were not primarily iconic representations of leaves, and these shapes were not documented as culture-specific pictorial conventions for leaves either. Yet, the speaker does produce a verbal description that overlaps with the depiction phrase. The source of evidence was therefore coded as “co-depiction speech.”

Using this evidence systematically for analyzing both gestural and depiction units where there was some degree of iconicity, we made the *operational* decision to code the dominant ground as Iconic only when the evidence was (a) primary

iconicity, i.e., when the representation was transparent enough to be interpreted even in the absence of (b) and (c). In the latter two cases, the semiotic ground was coded as Symbolic (i.e., conventional). For a gestural or depiction segment to be coded as Indexical, its function would need to be to draw attention to an element of the communicative context (e.g., a drawn arrow pointing at a specific part of the drawing). Using this method of analysis, we could establish some characteristics of polysemiosis in the two cultural practices, as described in the following two sections.

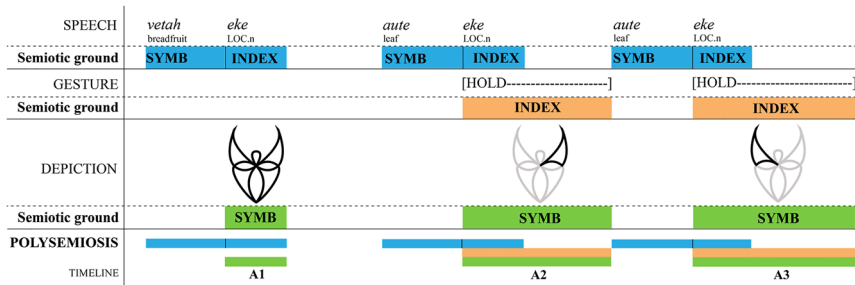
## 5 Paamese sand drawing

On Paama, one of the few islands where the Vanuatu cultural practice of sand drawing is still performed, it is quite prestigious and draws crowds of people at local culture festivals. Sand drawing is known as *mutis en atan* ('drawing in the ground') in Paamese, and is typically performed by men, but not exclusively. In contrast to the sand storytelling tradition of Central Australia, not all Paamese sand drawings have a core narrative function. They can be used to convey traditional knowledge about the local flora and fauna to the next generation, or used as part of rituals to summon spirits, seal a marriage agreement, remember historical landmarks, or even as keys to navigate the afterlife (Devylder 2022).

Another key dimension of Vanuatu sand drawing is that it constitutes intellectual property. This knowledge is indeed protected by traditional copyrights that have deep significance in Vanuatu culture, and sand drawing pattern ownership was exchanged as part of a pre-colonial Northern-Central Vanuatu network along with other goods like livestock, color pigments, and cultural items like bark belt, or pottery (Huffman 1996). Ownership of a sand drawing is tied to its place of origin; if it is performed somewhere else, a displacement ritual must be performed. The *talimbur* (Latin *cycas seemannii*) and *siel* (Latin *cordyline fruticosa*) leaves are planted in the ground to set a respectful perimeter around the drawing space. The green leaves are found on the Vanuatu coat of arms, and are associated with high status throughout the archipelago, hereby revealing the sacred, prestigious, and respected character of *mutis en atan* (Devylder 2022).

Once the perimeter is set (if needed), a Paamese sand drawing performance starts with cleaning the fine black volcanic ground with the palm of the hand or with a coconut leaf broom. The sand drawer then sprinkles some white ashes over the drawing space so as to strengthen the contrast between the black line and its white background. Sand drawers then use their index fingers and trace a grid through which their finger will draw a continuous line that takes an intricate geometrical path. Typically, the line should start and end at the same point and never take the same path twice. Paamese sand drawing patterns greatly varies in their geometrical





**Figure 8:** Polysemiotic utterance A, consisting of three segments A1, A2, A3.

complexity, from 13 to 229 strokes (Devyllder 2022). When the tracing of the line is done, the sand drawer will contemplate the drawing and begin a verbal description of its various parts and interpret some of its meaning to the audience.

Figure 8 is a simplified schematic representation of a relevant segment of a ‘breadfruit’ drawing, based on analyses like that in Figure 7. After having traced the geometrical pattern in the ground, the sand drawer starts the commentary part of the performance with the verbal utterance *vetah eke* (‘here is a breadfruit’). There is no primary iconicity between the drawing and an actual breadfruit, shown in Figure 9, making the semiotic ground predominantly symbolic. It is not surprising that the first *polysemiotic utterance*<sup>8</sup> encountered in the performance serves the goal of establishing a relation between the depiction and the speech systems by means of a semantic convention (*vetah* = breadfruit), especially since the audience is here another chief who also speaks Paamese. The speaker uses this shared language knowledge and anchors this meaning in the drawn expression with the help of a locative noun *eke* (‘here’) to link it to the first D-phrase (1a).

The interpretation of this first polysemiotic segment is thus something like: “the drawing that you see here, let’s agree that it represents a breadfruit, which I know you know what it is.” This whole meaning is much more compressed, yet as expressive thanks to the polysemiosis of the three sign systems. To spell out once more: the part “breadfruit, which I know you know what it is” is compressed in the sign *vetah*, and this is linked to the drawing with the help of another indexical sign [*eke*/ + ‘here’], thus forging the depiction phrase made of the previously produced geometrical tracing in the ground.

In segments A2 and A3, the sand drawer elaborates the whole meaning by singling out parts of the whole drawing. Once again, he uses a combination of the

<sup>8</sup> Following Kendon (2004) we extend the meaning of the term “utterance” from monosemiotic and unimodal speech to polysemiotic and multimodal.





Figure 9: The *vetah* sand drawing and a photograph of a breadfruit.

indexical linguistic sign *eke* supplemented by an indexical gesture, which indicates with more precision than speech what parts of the drawing is singled out exactly (i.e., the arcs in bold). Using a similar strategy as in A1 the sand drawer combines the meaning of the conventional linguistic sign [/aute/ + ‘leaf’] with indexical speech and gesture to produce the composite polysemiotic sign (LEAF). The overall communicative function of polysemiotic utterance A can thus be said to be that of *thematizing*: establishing the nature of the main topic of the polysemiotic discourse.

Polysemiotic utterance A is immediately followed by polysemiotic utterance B, shown schematically in Figure 10 (See also Figure 1). Having established the theme/topic in the previous polysemiotic utterance, here the sand drawer engages in a *commenting* function. Using combinations of signs from the different sign systems, with different semiotic grounds, he expresses the fact that the leaves of the breadfruit grow upwards. He makes a two-hand stroke-hold gesture with his two index fingers departing from the two leaves of the drawing and going upwards on the horizontal plane, while simultaneously vocalizing the iconic ideophone *clik-clik-clik* that sequentially goes up in pitch with the upward stroke-hold gesture in B1. The post-stroke HOLD part of the gesture nucleus – where the speaker keeps

SPEECH	<i>clik-clik-clik</i> <small>ideophone</small>	<i>ke</i> <small>LOC.n</small>	<i>keie</i> <small>PRO.3SG</small>	<i>sooni</i> <small>develop.leaves</small>	<i>vinaa</i> <small>go.up</small>	<i>nesa</i> <small>LOC.n</small>
Semiotic ground	ICON	INDEX	SYMB	SYMB		
GESTURE	[STROKE----HOLD-----]			[STROKE----HOLD-----]		
Semiotic ground	ICON			ICON		
POLYSEMIOSIS TIMELINE	B1		B2		B3	

Figure 10: Polysemiotic utterance B, consisting of three segments B1, B2, and B3.

his hands up in the air – overlaps with speech *ke keie* ('here it') in B2. The gesture has dominant iconic ground as the evidence for its interpretation was primary iconicity. It briefly retracts to a place of rest and is immediately followed by a similar stroke-hold gesture that overlaps with the speech unit *sooni vinaa nesa* ('develop leaves upwards'). The ground is here likewise iconic, as the interpretation of the gesture can be made independently of the co-gesture speech.

In this manner, the sand drawer elaborates his explanation of the nature of breadfruits by anchoring the invisible height, up to where the leaves grow with a combination of an iconic hold gesture and the indexical linguistic sign *ke* in B2. Perhaps to clarify, in B3 he reorganizes the distribution of semiotic grounds over sign systems by harnessing the iconicity of the upward gesture in combination with a more semiotically fine-grained speech unit that consists of the serial verb construction *sooni vinaa nesa* ('develop leaves goes up upwards').

Several generalizations can be made on the basis of these analyses. The first is that there is both overlap and complementarity in the semiotic grounds of the different parallel systems in each polysemiotic utterance. A second is that indexical expressions, in speech and gesture, are essential for interlinking the signs in the different sign systems, and for formulating polysemiotic predications of the type "THIS is a leaf," "HERE it grows." Finally, given the thematizing-commenting structure of the two polysemiotic utterances A and B, it is characteristic that the latter provides more new information, expressed both verbally (with symbolic ground), and gesturally (with iconic ground). We can keep these in mind as we consider examples from a rather different polysemiotic communication practice in the following sub-section.

## 6 Pitjantjatjara sand storytelling

A Pitjantjatjara storyteller typically sits on the ground and clears a drawing surface in the sand by smoothing it and removing small stones and debris. This surface is often cleared at several points throughout the story and wiped again at its conclusion (see Section 3.2.1). Sand is ubiquitous in these communities and common locations include the yards outside people's homes, community ovals, dry creek beds, and sand dunes. Marks are typically made using a bent wire or stick called *milpa* ('story wire/stick') – the act of sand story telling is referred to as *milpatjunanyi* (literally 'placing the story stick') – or with the hand. Leaves, sticks and other objects can also be placed to mark characters or objects within the story, thus constituting a semiotic resource that is distinct from both gesture and depiction, and awaits further study (Green 2014).

The performance is inherently multimodal *and* polysemiotic, understanding these as complementary and anything but synonymous notions (see Section 2) including (a) depictions created in the sand, (b) gestures (distinct from the movements to create the depictions), (c) the sounds of the stick hitting the sand and of course, and (d) the spoken utterances. Some of the depictions are culturally formalized, relating to conventions also seen in various other art forms, while others are spontaneous expressions with iconic or indexical ground (Green 2014). The wire is often used to beat rhythmically throughout the storytelling. This provides an accompaniment and emphasis to parts of the story as well as being a distinct marker of this genre of verbal art, as discussed in Section 2.2.2.<sup>9</sup> Women and girls often carry *milpa* with them throughout the day and *milpatjunanyi* is seen throughout the community several times each day.

The Pitjantjatjara sand stories analyzed within this project were recorded in Pukatja, also known as Ernabella, a community of approximately 500 people in the Anangu Pitjantjatjara Yankunytjatjara lands in the north-west corner of South Australia. Comparing sand stories as told by Warlpiri and Pitjantjatjara women, Munn (1973) noted that the Pitjantjatjara sand stories appeared to have a looser relationship between depiction and speech with many marks viewed as just marks either as a kind of doodling or the end-points of gestures.<sup>10</sup> There are, however, also many examples where the depiction, speech and gesture are tightly integrated, such as those discussed below.









The particular sand story analyzed here is told by a woman to her family and the camera. It is one of the common “back in my day” story themes and she describes how, when she was young, she and the other children would go out to play in the bush after school. The following two polysemiotic utterances immediately follow each other and are taken nearly halfway through the story as she is talking about how they would catch donkeys and then ride them. Each example goes from one complete wipe to the next, thus covering a whole D-frame (see Section 3.2.1).

Figure 11 is a schematic representation of a polysemiotic utterance, with segments C1, C2, C3, C4, and C5. C3 and C4 could possibly be combined as they are produced as a single fluent articulation arc and a single clause of speech; however, we have separated them due to the different utilizations of polysemiosis. Figure 12 shows the final state of the sand drawing before it is wiped.

---

<sup>9</sup> The wire is also often beaten without specific intention, almost absentmindedly. One woman commented that beating the wire “gives the girls space to gather their thoughts, give their mind a rest” (Eickelkamp 2011: 32).

<sup>10</sup> See Green (2014) for a discussion of some sand drawing marks as the visible traces of gesture in Arandic sand stories.

SPEECH		<i>ka</i> and.DS	<i>tanki</i> donkey	<i>tjuja</i> PL	<i>ngugurari-pai</i> be in between- HAB	<i>ka-la</i> and.DS-we	<i>angajjura</i> blocking	<i>witil-pai</i> catch-HAB	<i>munu-la</i> and.SS-we	<i>tatil-pai</i> ride-HAB	
Semiotic ground		SYMB									
GESTURE		[STROKE... J									
Semiotic ground		INDEX									
DEPICTION											
Semiotic ground	ICONIC	SYMBOLIC					SYMB	SYMB		SYMB	
POLYSEMIOSIS	[Timeline bar]										
TIMELINE	C1	C2					C3	C4		C5	

**Figure 11:** Polysemiotic utterance C, consisting of four segments C1, C2, C3, C4, and C5.

In segment C1, the storyteller draws an arc in the sand which depicts a barrier in the landscape to help set the scene for what follows. The semiotic ground of this expression is iconic, given that the evidence used was based on primary iconicity (there was neither co-depiction speech, nor convention to provide its meaning). It is a form of *diagrammatic* iconicity, since this shape is often used for creeks, roads, windbreaks, and other things which share this shape. It is interesting to note that the speaker uses productive depiction as a standalone *monosemiotic* segment, which is not contained elsewhere within the utterance.

Segment C2 combines speech and depiction. The speech starts with the different subject coordinator indicating a shift of subject from the previous clause.



**Figure 12:** A still frame showing the final state of the depiction frame in polysemiotic utterance C.

The storyteller says *ka tanki tjuta ngururari pai* ('and donkeys would be in the space between'). The semiotic ground was here judged to be symbolic, as the evidence was co-depiction speech (see Section 4). At the same time, the storyteller depicts the donkeys by tapping a series of dots into the sand near the afore placed barrier.

Segment C3 is very similar. As the storyteller produces the first part of the converbial clause *kala angatjura* ('and we by blocking ...'), she depicts the action of blocking off the donkeys with a sweeping movement of the wire. The semiotic ground of both speech and depiction here was coded as (predominantly) symbolic, for the same reason as above.

One could possibly have analyzed the sweeping movement which leaves a mark, but also continues further along its trajectory lifting off the sand, as an iconic gesture, but we have not done so as we have aimed to maintain a distinction between the systems of gesture and depiction whenever possible, for conceptual and methodological reasons, as discussed in Section 2. So, in our analysis, gesture only occurs in C4 when the storyteller points to one of the 'donkey' dots produced in C2 as she says *witilpai* ('catch'), thus using indexicality to single out the object of the action described in speech. The depiction here shifts from produced to commented while the predominant ground was again coded as symbolic by inheritance from the production in C2. The speech is again symbolic.

Finally, in C5, the speaker uses speech and depiction, both with symbolic ground, to communicate that they rode the donkey. In speech, she uses the same subject coordinator *munu* and says *munula tatilpai* ('and we would ride') At the same time, she draws a single line in place of one of the already placed dots. This represents riding a donkey due to the symbolic convention in the community of using parallel lines to refer to animals which are being ridden. After this, the sand surface is wiped and a new D-frame is begun.

Polysemiotic utterance D immediately follows. It is divided into three polysemiotic segments of speech and depiction, with gesture appearing at the end. D1 features the speech *kutjara kutjarala tatilpai kutjara kutjara* ('we would ride two by two, two by two'), and conventionalized depictions of a donkey or horse with two people sitting astride it: two parallel lines representing the animal with two curved lines representing the people sitting on top. Thus, both sign systems rely on symbolism, and their meanings overlap, but are at least somewhat complementary, since the animal that is being ridden is only expressed in the depiction system.

D2 is very similar, with the depiction producing another donkey with two riders, while the speech diverges to focus on the specific location of the riding with an indexical demonstrative 'here on the donkey.' In D3, speech (*alatji* 'like this') and a pointing gesture indicate to the way in which the donkey is depicted. This is then a commented D-phrase with co-depiction speech. After this, the D-frame ends, as the whole drawing is again wiped.

SPEECH	<i>kujara kujara-la</i> two by two-we	<i>tatil-pai ride-HAB</i>	<i>kujara kujara</i> two by two	<i>nyangka</i> here	<i>tangkiyingka</i> on the donkey	<i>munu-la</i> and.SS-we	<i>tangkiy</i> i donkey	<i>ponya</i> ANAP H	<i>alajji</i> like this	<i>walkajunkun-pai</i> draw-HAB
Semiotic ground	SYMB		INDEX	SYMB			INDEX	SYMB		
GESTURE										[STROKE...]
Semiotic ground										INDEX
DEPICTION	/	//	///	////	/////	////	///	///	///	wipe
Semiotic ground	SYMB		SYMB			SYMB				
POLYSEMIOSIS	[Timeline bar]									
TIMELINE	D1			D2			D3			

Figure 13: Polysemiotic utterance D, consisting of four segments D1, D2, and D3.



Figure 14: A still frame showing the final state of the depiction frame in polysemiotic unit D.

Comparing these Pitjantjatjara examples with the Paamese examples in the previous section shows considerable similarities but also some significant differences. Returning to the three generalizations made at the end of that section, we can note that indeed, combining segments with two or three different (dominant) semiotic grounds appears to be the rule in polysemiotic utterances with sand drawing. Interestingly, in all examples, the depictions appear to be dominated not by iconicity but by symbolcity, since either co-depiction speech or prior conventions were required for interpreting the marks left in the sand. Only the first segment of utterance C was in fact coded with iconic depiction, and it is notable that this was in the Pitjantjatjara performances, which on the whole were much more spontaneous than those of the stylized Paamese sand drawings. At the same time, we should note that this conclusion was based on the operational decision to treat evidence (b) co-



gesture/depiction speech and (c) culture-specific convention, as equally strong in deciding to code the ground as symbolic (see Section 4). If we had limited this to (c), then the analysis of the depiction signs in utterance C would in fact have been predominantly iconic.

Concerning the key role of indexicality, the final polysemiotic segment D3 was remarkably similar to that of the Paamese segments A2 and A3, with both speech and gesture indexically relating the symbolic meaning of the previously drawn depiction with the symbolic meaning of the spoken words.<sup>11</sup> The Paamese segment B2, where only speech performs this binding function, is in a way “the exception that proves the rule,” since the performer is here engaged in a predominantly iconic gesture which, in a way, extends the depiction into the air.

Finally, there is clear topical continuity between the A and B, and the C and D polysemiotic utterances, with the second ones in these pairs building upon the former, in a type of theme-comment information structure. However, due to the occurrence of the Pitjantjatjara examples in the middle of the story, rather than at the beginning as in the Paamese pairs, and the fact that there was commenting being done in C as well, this similarity is only partial.

Turning to differences between the two cultural practices, the most obvious is one is the fact that most of Paamese Depiction phrases were *commented*, while in Pitjantjatjara these were mostly *produced* (see Section 3.2). This kind of depiction that takes place at the same time as the story develops, lends itself to a more “gesture-like” depiction where not only the marks on the sand, but also the motions used to create are expressive. Gestures which are wholly outside of the sand space as in the Paamese polysemiotic utterance B do occur in *milpatjunanyi*, but are fairly rare, at least in the Pitjantjatjara sand stories that we analyzed. This contributes to some degree of blurring between the systems of depiction and gesture as a movement occurs partly in the sand and partly in the air, such as the blocking depiction in C3. As noted there, however, we did not code the gestural aspect in this particular case, and in general, so as to focus on the properties and meanings of the individual sign systems. Indeed, this was fruitful for example in C1, where depiction occurred monosemiotically, which is the only segment among the examples discussed in this section where the dominant semiotic ground of depiction is (clearly) iconic. The basic independence of depiction from gesture is quite evident in example D schematized in Figure 13 and with the final frame shown in Figure 14.

---

<sup>11</sup> This is further expanded by the use of the anaphoric marker *panya* so that the Pitjantjatjara speech *tangkiyi panya* could be translated as ‘the donkey that I know that you know.’

## 7 Summary and conclusions

The aim of the research reported in this article has been threefold. First, we needed to define theoretically the key notions involved in polysemiosis (or polysemiotic communication), including signs, signals and the three universal human sign systems of language, gesture and depiction, with their corresponding material and semiotic properties. While some of these have been discussed in previous research (e.g., Zlatev 2019; Zlatev et al. 2020), the presentation in Section 2 provides the most explicit presentation of the cognitive-semiotic framework of polysemiosis so far.

Second, departing from this framework, we formulated a coding system implemented in the software ELAN, and to use this in the systematic analysis of a polysemiotic and multimodal corpus of 23 Paamese sand drawings and 20 Pitjantjatjara sand stories, as described in Section 3. In doing so, we did our best to present the criteria for identifying segments in the three systems as explicitly and reliably as possible, acknowledging some (but not too much) hierarchical structure. The aim was to make our analysis both practical and, in principle, replicable. It is our experience that much valuable theoretical work within the fields of “multimodality” and semiotic analysis (e.g., O’Halloran et al. 2016) could benefit from such explicit operational definitions in this respect.

The third aim was to “test” the system in practice, and in Section 4, we illustrated how the analysis can be applied to a fragment of polysemiotic communication in the Paamese data, and stated some more operational definitions, in particular on how to decide the *dominant semiotic ground* for every specific sign, which is a well-known problem in semiotics (e.g., Jakobson 1965). In the case of language/speech the fact that this ground is conventional (and hence to be coded as symbolic) is uncontroversial and the decision to mark this as such, with the exception of ideophones and demonstratives, was methodologically unproblematic. But for gesture and depiction, it is in fact both controversial (e.g., Kendon 2004; Sonesson 1989; Zlatev 2015b) and methodologically hard to decide. Even when there was a clear iconic ground in the *diagrammatic representations* of gestural and depictive signs, this most often also coincided with (degrees of) conventionality, derived either from prior conventions, or from co-gesture/co-depiction speech, which is typical for diagrams, as pointed out by Peirce and acknowledged by Jakobson (1965).

However, we decided to be quite conservative in our coding, and marked even gestural and depiction units with noticeable iconicity as symbolic if one could not interpret them on the basis of primary iconicity (Sonesson 1997). On the one hand, this allowed us to compare the data from the two different practices systematically, as shown in Sections 5 and 6, and to make some generalizations about similarities and differences in how they utilize polysemiosis. On the other hand, by glossing over



the distinction: “prior convention/co-depiction speech,” we may have been *too* conservative, thus failing to account for the (intuitively) higher degree of iconicity in the Pitjantjatjara utterance C than in D (see Section 6). But this is something that can be easily tested and if needed rectified in further analysis, since once the evidence for coding for the corpus has been consistently made, one can easily change the operational definition. One could, for example, regard the primary-iconicity/convention as a cline, and treat examples such as Pitjantjatjara C as being on the middle.

With respect to making a systematic distinction between gesture and depiction, our approach was different from that of Green (2014), with whom we are otherwise mostly in agreement. This was a matter of principle for the theory, since in general, the two sign systems are independent of one another: one can gesture without drawing and draw without gesturing. And both are independent from language/speech which can very well occur on its own.

On the other hand, empirical examples such as “wire-tapping,” or the Pitjantjatjara utterance C analyzed in Section 6, show that it is indeed hard to distinguish the bodily movement itself, from the trace that it makes on the ground. But once again, the coding system used was sufficiently versatile to capture such overlaps, which parallel units in both systems. Importantly, the criteria for both need to be consistent, which was in fact a feature of our operationalization. On the other hand, operationalizations (and theories) are always simplifications of the phenomenon studied. Hence, only future research will show to what extent the decisions of the present analysis will be maintained or would need to be modified. After all, this is the consequence of being true to the conceptual-empirical loop of cognitive semiotics.

**Acknowledgements:** We would above all wish to thank the participants on Paama and in the Anangu Pitjantjatjara Lands. In addition, we thank Heron Kulukul, Ezekiel Toka, and Lucie Laporte-Devylder for assistance with transcription of the Paamese data and Sasha Wilmoth for help with the Pitjantjatjara data. The research was supported by grant 2018–01281 from the Swedish Research Council.

## References

- Ahlner, Felix & Jordan Zlatev. 2010. Cross-modal iconicity: A cognitive semiotic approach to sound symbolism. *Sign System Studies* 38. 298–348.
- Andrén, Mats. 2010. *Children's gestures between 18 and 30 months*. Lund: Media Tryck.
- Arbib, Michael. 2005. From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics. *Behavioral and Brain Sciences* 28. 105–168.
- Baron, Jacopo. 2020. *Gardens of the mind: A study on Vanuatu sand-drawing*. Paris: EHESS PhD dissertation.
- Burling, Robbins. 2005. *The talking ape*. Oxford: Oxford University Press.
- Chandler, Daniel. 1994. *Semiotics: The basics*. London: Routledge.

- Clark, Herbert H. 1996. *Using language*. Cambridge: Cambridge University Press.
- Coseriu, Eugenio. 1985. Linguistic competence: What is it really? *The Modern Language Review* 80(4). xxv–xxxv.
- Coseriu, Eugenio. 2000. The principles of linguistics as a cultural science. *Transylvanian Review (Cluj)* 9(1). 108–115.
- Deacon, A. Bernard & Camilla H. Wedgwood. 1934. Geometrical drawings from Malekula and other islands of the New Hebrides. *The Journal of the Royal Anthropological Institute of Great Britain and Ireland* 64. 129–175.
- Devylder, Simon. 2018. Diagrammatic iconicity explains asymmetries in possessive constructions: Evidence from Paamese. *Cognitive Linguistics* 29(2). 313–348.
- Devylder, Simon. 2019a. Black sand stories: A polysemiotic and multimodal documentation of Paamese sand stories, a critically endangered tradition of Vanuatu. *Endangered Languages Archive*. <https://www.elararchive.org/dk0565/> (accessed 4 May 2023).
- Devylder, Simon. 2019b. Review of Dancygier. In Barbara (ed.), *The Cambridge handbook of cognitive linguistics. English text construction*, vol. 12(1), 143–153. Cambridge, UK: Cambridge University Press.
- Devylder, Simon. 2022. The archipelago of meaning: Methodological contributions to the study of Vanuatu sand drawing. *The Australian Journal of Anthropology* 33(2). 279–327.
- Diessel, Holger. 2006. Demonstratives, joint attention, and the emergence of grammar. *Cognitive Linguistics* 17(4). 463–489.
- Diget, Ida S. K. 2019. *Intersemiotic translation from film to audio description: A cognitive semiotic approach*. Lund: Lund University Master's thesis.
- Dingemanse, Mark. 2012. Advances in the cross-linguistic study of ideophones. *Language and Linguistics Compass* 6. 654–672.
- Donald, Merlin. 1991. *Origins of the modern mind: Three stages in the evolution of human culture*. Cambridge, MA: Harvard University Press.
- Eickelkamp, Ute. 2011. Sand storytelling. In Ute Eickelkamp (ed.), *Growing up in Central Australia: New anthropological studies of aboriginal childhood and adolescence*, 103–130. New York: Berghahn.
- Forceville, Charles. 2017. Visual and multimodal metaphor in advertising. *Styles of Communication* 9(2). 26–41.
- Gell, Alfred. 1998. *Art and agency: An anthropological theory*. Oxford: Clarendon Press.
- Green, Jennifer. 2014. *Drawn from the ground: Sound, sign and inscription in Central Australian sand stories*. Cambridge: Cambridge University Press.
- Hockett, Charles F. 1960. The origin of speech. *Scientific American* 203. 89–97.
- Huffman, Kirk. 1996. Trading, cultural exchange and copyright: Important aspects of Vanuatu arts. In Joël Bonnemaïson, Kirk Huffman, Christian Kaufmann & Darrel Tryon (eds.), *Arts of Vanuatu*, 182–194. Honolulu: University of Hawai'i Press.
- Husserl, Edmund. 1989. *Ideas pertaining to a pure phenomenology and to a phenomenological philosophy* (Studies in the phenomenology of constitution 2). Dordrecht: Kluwer.
- Itkonen, Esa. 2003. *What is language? A study in the philosophy of linguistics*. Turku: Turku University Press.
- Jakobson, Roman. 1965. Quest for the essence of language. *Diogenes* 13. 21–38.
- Kendon, Adam. 2004. *Gesture: Visible action as utterance*. Cambridge: Cambridge University Press.
- Kravchenko, Alexander. 2017. Making sense of languaging as a consensual domain of interactions: Didactic implications. *Intellectica* 2(68). 175–191.
- Layard, John. 1936. *Stone men of Malekula: Vao*. London: Chatto and Windus.
- Louhema, Karoliina, Jordan Zlatev, Maria Graziano & Joost van de Weijer. 2019. Translating from monosemiotic to polysemiotic narratives: A study of Finnish speech and gestures. *Sign Systems Studies* 47(3/4). 480–525.

- McGurk, Harry & John MacDonald. 1976. Hearing lips and seeing voices. *Nature* 264. 746–748.
- Müller, Cornelia. 2016. From mimesis to meaning: A systematics of gestural mimesis for concrete and abstract referential gestures. In Jordan Zlatev, Goran Sonesson & Piotr Koderak (eds.), *Meaning, mind and communication: Explorations in cognitive semiotics*, 211–226. Frankfurt am Main: Peter Lang.
- Müller, Cornelia. 2017. Waking metaphors: Embodied cognition in multimodal discourse. In Beate Hampe (ed.), *Metaphor: Embodied cognition and discourse*, 297–316. Cambridge: Cambridge University Press.
- Munn, Nancy D. 1973. *Walbiri iconography: Graphic representation and cultural symbolism in a Central Australian society*. Ithaca & London: Cornell University Press.
- O'Halloran, Kay L., Sabine Tan & Peter Wignell. 2016. Intersemiotic translation as resemiotization: A multimodal perspective. *Signata* 7(3). 199–229.
- Patterson, Mary. 2006. Agency, kinship, and history in North Ambrym. *Journal of the Royal Anthropological Institute* 12(1). 211–217.
- Rio, Knut. 2005. Discussions around a sand-drawing: Creations of agency and society in Melanesia. *Journal of the Royal Anthropological Institute* 11(3). 401–423.
- Ryan, Marie-Laure. 2007. Toward a definition of narrative. In David Herman (ed.), *The Cambridge companion to narrative*, 22–35. New York: Cambridge University Press.
- Ryan, Marie-Laure. 2012. Narration in various media. In Peter Hühn, John Pier, Wolf Schmid & Jörg Schönert (eds.), *The living handbook of narratology*. Hamburg: Hamburg University.
- Saeed, John. 2016. *Semantics*, 4th edn. Oxford: Blackwell.
- Sahlgren, Magnus. 2006. *The word-space model: Using distributional analysis to represent syntagmatic and paradigmatic relations between words in high-dimensional vector spaces*. Stockholm: Stockholm University PhD dissertation.
- Sandler, Wendy. 2012. Dedicated gestures, and the emergence of sign language. *Gesture* 12. 265–307.
- Saussure, Ferdinand de. 1916. *Cours de linguistique générale*. Paris & Lausanne: Payot.
- Savage-Rumbaugh, Sue & Richard Lewin. 1994. *Kanzi: The ape at the brink of the human mind*. New York: Wiley.
- Sonesson, Göran. 1989. *Pictorial concepts: Inquiries into the semiotic heritage and its relevance for the analysis of the visual world*. Lund: Aris/Lund University Press.
- Sonesson, Göran. 1997. The ecological foundations of iconicity. In Irmengard Rauch & Gerald F. Carr (eds.), *Semiotics around the world: Synthesis in diversity*, 739–742. Berlin: Mouton de Gruyter.
- Sonesson, Göran. 2007. From the meaning of embodiment to the embodiment of meaning: A study in phenomenological semiotics. In Tom Ziemke, Jordan Zlatev & Roslyn Frank (eds.), *Embodiment* (Body, language, and mind 1), 85–128. Berlin: Mouton de Gruyter.
- Sonesson, Göran. 2010. From mimicry to mime by way of mimesis: Reflections on a general theory of iconicity. *Sign Systems Studies* 38(1/4). 18–65.
- Stampoulidis, Georgios. 2019. Stories of resistance in Greek street art: A cognitive-semiotic approach. *The Public Journal of Semiotics* 8(2). 29–48.
- Stampoulidis, Georgios. 2021. *Street activism on Athenian walls: A cognitive semiotic analysis of metaphor and narrative in street art*. Lund: Lund University PhD dissertation.
- Tamura, Shinichi & Shingo Kawasaki. 1988. Recognition of sign language motion images. *Pattern Recognition* 21(4). 343–353.
- Taylor, John. 2005. Paths of relationship, spirals of exchange: Imag(in)ing North Pentecost kinship. *Australian Journal of Anthropology* 16(1). 76–94.
- Tjitayi, Katrina & Sandra Lewis. 2011. Envisioning lives at Ernabella. In Ute Eickelkamp (ed.), *Growing up in Central Australia: New anthropological studies of aboriginal childhood and adolescence*, 49–62. New York: Berghahn.
- UNESCO. 2006. *Masterpieces of the oral and intangible heritage of humanity*. Paris: UNESCO.

- Vandelanotte, Lieven. 2017. Viewpoint. In Barbara Dancygier (ed.), *The Cambridge handbook of cognitive linguistics*, 157–171. Cambridge: Cambridge University Press.
- Wittenburg, Peter, Hennie Brugman, Albert Russel, Alex Klassmann & Han Sloetjes. 2006. ELAN: A professional framework for multimodality research. Paper presented at the *Fifth international conference on language resources and evaluation* (LREC 2006).
- Wittgenstein, Ludwig. 1953. *Philosophical investigations*. Oxford: Blackwell.
- Zlatev, Jordan. 1997. *Situated embodiment: Studies in the emergence of spatial meaning*. Stockholm: Gotab.
- Zlatev, Jordan. 2014. Human uniqueness, bodily mimesis and the evolution of language. *Humana* 7(27). 197–219.
- Zlatev, Jordan. 2015a. Cognitive semiotics. In Peter Trifonas (ed.), *International handbook of semiotics*, 1043–1067. Dordrecht: Springer.
- Zlatev, Jordan. 2015b. The emergence of gestures. In Brian MacWhinney & William O’Grady (eds.), *The handbook of language emergence*, 458–477. New York: Wiley.
- Zlatev, Jordan. 2019. Mimesis theory, learning, and polysemiotic communication. In Michael Peters (ed.), *Encyclopedia of educational philosophy and theory*. Dordrecht: Springer.
- Zlatev, Jordan, Göran Sonesson & Piotr Konderak. 2016. *Meaning, mind, and communication: Explorations in cognitive semiotics*. Frankfurt am Main: Peter Lang.
- Zlatev, Jordan, Marta Sibierska, Przemysław Zywiczyński, Monika Boruta-Zywiczyńska & Joost van de Weijer. in press. Can pantomime narrate? A cognitive-semiotic approach. In Przemysław Zywiczyński & Johan Blomberg (eds.), *Perspectives on pantomime*. Amsterdam: Benjamins.
- Zlatev, Jordan & Mats Andrén. 2009. Stages and transitions in children’s semiotic development. In Jordan Zlatev, Mats Andrén, Marlene Johansson-Falck & Carita Lundmark (eds.), *Studies in language and cognition*, 380–401. Newcastle: Cambridge Scholars.
- Zlatev, Jordan, Przemysław Żywiczyński & Sławomir Waciewicz. 2020. Pantomime as the original human-specific communicative system. *Journal of Language Evolution* 5(2). 156–174.