

# Translating from monosemiotic to polysemiotic narratives: A study of Finnish speech and gestures

Karoliina Louhema<sup>1</sup>, Jordan Zlatev<sup>2</sup>, Maria Graziano<sup>3</sup>,  
Joost van de Weijer<sup>4</sup>

**Abstract.** Human communication can be either monosemiotic or polysemiotic, depending on whether it combines ensembles of representations from one or more *semiotic systems* such as language, gesture and depiction. Each semiotic system has its unique storytelling potentials, which makes intersemiotic translation from one system to another challenging. We investigated the influence of the source semiotic system, realised in speech and a sequence of pictures, respectively, on the way the same story was retold using speech and co-speech gestures. The story was the content of the picture book *Frog, Where Are You?*. A group of Finnish speakers saw the story in pictures, and another group heard it in matched oral narration. Each participant retold the story to an addressee and all narrations were video-recorded and analysed for both speech and gestures. Given the high degree of iconicity in depiction, we expected more iconic gestures (especially enactments) in the narratives translated from pictures than in those translated from speech. Conversely, we expected greater narrative coherence in the narratives translated from speech. The results showed that more iconic gestures were produced in the narratives translated from speech, but these were primarily not from the enactment subtype. As expected, iconic enactments were more frequent in the narratives translated from the story presented in pictures. The narratives produced by participants who had only heard the story did not have a greater variety of connective devices, yet the type of devices differed slightly between the groups. Together with some additional differences between the groups that had not been anticipated, the results indicate that a story presented in different semiotic systems tends to be translated into different polysemiotic narratives.

**Keywords:** polysemiotic communication; intersemiotic translation; multimodality; cognitive semiotics; iconicity

---

<sup>1</sup> Centre for Languages and Literature, Lund University, Sweden; e-mail: karoliina.louhema@gmail.com.

<sup>2</sup> Centre for Languages and Literature, Lund University, Sweden; Centre for Language Evolution Studies, Nicolaus Copernicus University, Torun, Poland; e-mail: jordan.zlatev@semiotik.lu.se.

<sup>3</sup> Lund University Humanities Lab, Sweden; e-mail: maria.graziano@humlab.lu.se.

<sup>4</sup> Centre for Languages and Literature, Lund University; Lund University Humanities Lab, Sweden; e-mail: joost.van\_de\_weijer@ling.lu.se.

## 1. Introduction

Herbert H. Clark (2004: 457) formulated the task of “telling a story” in the following terms:

Telling stories is hard. First you need something to tell – a genuine *story*. You must be acquainted with a happening, or series of events, about which you can make a point. [...] Even before you start, you need an overall plan – where to begin, what events and evaluations to include or emphasize, and where to end. In the actual telling, you need to *formulate* utterances one by one. For each utterance, you must select the right *words and gestures from the repertoire available* to you – your *language and culture* – and you must do so in a timely fashion. (our emphasis)

This quote captures several features that are central for the topic of this paper. First, a story involves a series of events, with a beginning, middle and end. Second, when the story is *narrated*, these events need to be ordered and evaluated by the storyteller, and adapted to the audience, as well as to the shared culture in general. Third, the narrative will necessarily be expressed in one or more *semiotic systems*: the most important ones being those of *language, gesture, and depiction*, with many subdivisions based on the media and sensory modalities involved (Zlatev 2019; Stampoulidis, Bolognesi, Zlatev 2019). Most human communication involves a combination of two or more semiotic systems and is thus *polysemiotic*. As we explain in Section 2, the notion of polysemiotic communication has advantages over the more broadly known term *multimodality* (Kress 2010).

Each semiotic system has its unique storytelling potentials, which makes the *intersemiotic translation* (Jakobson 1959) of a given story from one system to another a challenging task. What is particularly understudied is how the same story, expressed in either speech or pictures, is translated into a polysemiotic narrative consisting of speech and gestures. Given the properties of the source semiotic systems that we discuss in Section 2, it could be expected that hearing the story would result in more coherent narratives in terms of organization and plot development. On the other hand, seeing the story in pictures may give rise to more perceptually detailed narratives, including iconic (i.e. resemblance-based) gestures.

Using a cognitive-semiotic framework, we here explore these issues first theoretically and then empirically by comparing two groups of Finnish participants exposed to the same story in two different systems (speech vs. depiction), and asking each participant to “retell” it to an addressee in a spontaneous narration. By studying the resulting narratives qualitatively and quantitatively, we aim to contribute to narratology, since pictorial and polysemiotic narratives

have been rather understudied (with some notable exceptions, e.g., Hühn *et al.* 2009; Ranta 2011; Sonesson 2015a; Yiheng Zhao 2015). In addition, we aim to gain a better understanding of the relations between the semiotic systems of language and gesture (Kendon 2004; Zlatev 2015). Empirically, the study contributes with data on polysemiotic narration and gesture in Finnish, where the topic has hardly been studied at all, except in the context of communicative difficulties (Jääskeläinen 2009; Haddington, Kääntä 2011).

The paper is divided in five sections. The theoretical background is presented in Section 2, which builds up the framework of the study. The aims of the study and research questions are taken up once again at the end of the section with this framework in place. The methodology and design of the study, together with a number of specific hypotheses are explained in Section 3. Section 4 presents the results, Section 5 the discussion, and finally, conclusions are given in Section 6.

## 2. The cognitive semiotic framework

Here, we introduce our cognitive semiotic framework, and the most relevant concepts for the present research, including our understanding of the notions of *sign* and *semiotic system*, with a review of specific properties of the systems of language, gesture and depiction. We further discuss some theoretical and terminological issues in narratology and translation theory as a backdrop to our research questions and general hypotheses presented at the end of the section.

### 2.1. Basic notions

Cognitive semiotics emerged over the past two decades to accommodate the growing need for a better understanding of the relations between mind, meaning and communication (Zlatev 2012; Sonesson 2015a; Konderak 2016; Lenninger 2016). Dedicated to the transdisciplinary study of meaning, cognitive semiotics combines concepts and methods from semiotics, linguistics and cognitive science. Researchers aim to integrate theoretical and empirical research favouring the use of a particular type of methodological triangulation, using the combination of 1st person (e.g. intuition), 2nd person (e.g. empathy) and 3rd person (e.g. experimentation) methods (Zlatev 2009). In this and other respects, cognitive semiotics is clearly influenced by phenomenology, especially the work of Edmund Husserl and Maurice Merleau-Ponty.

One of the principal concepts in cognitive semiotic research is that of the *sign*, with varying definitions in different cognitive semiotic theories. Avoiding both

too narrow a conception, as in Saussurean semiotics, and a too broad one, as in Peircean semiotics, but with closer affinity to the latter, Sonesson (2010) defines the sign as a socio-cognitive process where *expression* and *content* are both related and differentiated, with the former more directly perceived, whereas the latter is more in focus. Thus, the expression is something that can bring something other than itself to mind, implying the necessity of a *conscious interpreter*, as shown in Fig. 1, and not just an interpretant, as in Peircean theory (see Zlatev 2009; Ahlner, Zlatev 2010). It is this interpreter who needs to perceive the relationship between expression and content or, in other words, the *ground* for the sign. This ground can be of three kinds: *iconic* – based on similarity; *indexical* – based on association (e.g. spatiotemporal or a part-whole relation); or *symbolic* – based on convention, along the interpretation of the well-known Peircean trichotomy proposed by Sonesson (2010). Any specific sign (process) usually combines different grounds, and it is the *predominant* one that determines its type, as an iconic, indexical or symbolic sign (Jakobson 1965).

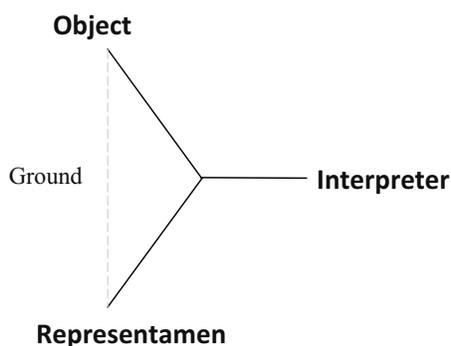


Figure 1. The relationship between *representamen* (expression), *object* (the content that the expression stands for), and *the interpreter*.

Iconic signs involve (combinations of) *primary* and *secondary* iconicity, illustrated in Fig. 2. In the former, the perception of similarity between expression and content is the ground for understanding the sign. In the latter, one must first know what the expression represents for the resemblance to be seen (Sonesson 1997, 2010). The second of the three grounds – indexicality – is based on time/space *contiguity* or *factorality* (part-whole relation) between two entities: as smoke is associated to fire, or a head is part of the body. Other examples of indexical signs could be a knock on the door as an index of someone at the door (Sonesson

2015b) or the typical performative index: the pointing gesture (Andrén 2010). Lastly, in symbolic signs the predominant ground is conventionality, an agreement between sign users. Sonesson (2010) mentions traffic regulations as an example of symbolic signs, and Zlatev (2009) emblematic gestures (e.g. the OK gesture), words and grammatical constructions.

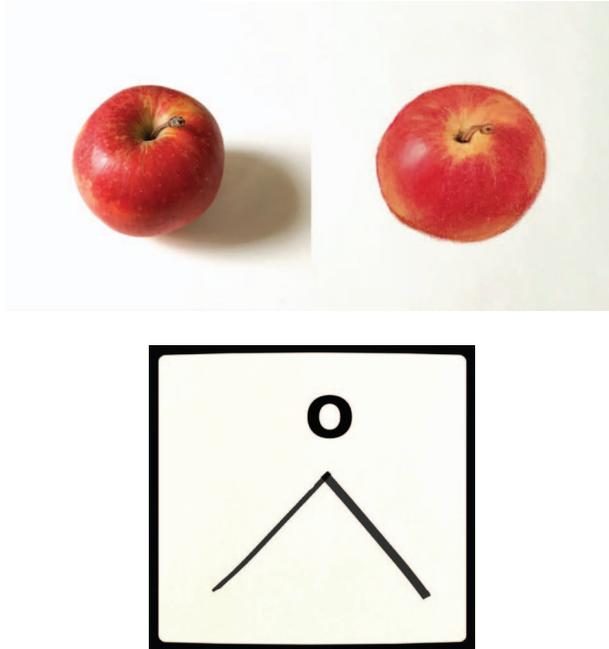


Figure 2. An image of an apple, with predominant primary iconicity, and a doodle with predominant secondary iconicity: is this the sun over a tent or a mountain, or a simplistic human figure?

## 2.2. Semiotic systems and polysemiotic communication

Signs form interrelations with other signs to form semiotic systems, which differ from one another in terms of semiotic properties (such as the dominant ground) and the features of the medium used for expression. Language, gesture and depiction are three universal human semiotic systems (Zlatev 2019), with some distinguishing properties, discussed in the following sub-sections, and shown in Table 1.

Table 1. The three universal semiotic systems of language, gesture and depiction, with some of their properties. (Adapted from Stampoulidis, Bolognesi, Zlatev 2019.)

Properties	Semiotic systems		
	Language (Speech, Writing, Signing)	Gesture	Depiction
<b>Production</b>	Vocal, Material, Body	Body	Material
<b>Perception (dominant)</b>	Auditory, Visual, Visual	Visual	Visual
<b>Permanence</b>	Low, High, Low	Low	Medium
<b>Double articulation</b>	Yes, Yes, Yes (?)	No	No
<b>Dominant semiotic ground</b>	Conventionality	Iconicity / Indexicality	Iconicity
<b>Syntagmatic relations</b>	Compositional	Sequential	Possibly sequential

### 2.2.1. Language

As shown in Table 1, language could be viewed as a superordinate semiotic system, which can be realized as spoken language, written language and signed language, as each of these has specific expressive potentials, and involves different sensory modalities (Green 2014). As such, language is predominantly (for there is also iconicity and indexicality involved, especially in signed languages) a system of *symbolic* signs, and could be defined as a predominantly “conventional-normative semiotic system for communication and thought” (Zlatev 2008: 37):

It is *conventional*, in the sense that it is based on an implicit or explicit agreement among its users. Even when implicit, these conventions are *normative*, in the sense of prescribing criteria for correct use. These conventions/norms are *semiotic*, since they take the form of signs [...], organized in a system. And finally this system is used not only for *communication* (through speech, signing, or writing) but also in *thinking*, after being internalized by children.

Being conventional does not imply being “arbitrary” (Ahlner, Zlatev 2010). Iconicity on the level of single spoken signs is reflected in *sound symbolism*, of which the most familiar kind are *onomatopoeic* expressions like ‘bang’ and ‘splash’. While individual speech sounds can thus be in part meaningful, a distinguishing feature for spoken language is that they usually function as *phonemes*, which are not signs themselves, but rather serve to distinguish lexical signs (words) from

one another. This feature is referred to as *duality of patterning* (Hockett 1960) or *double articulation* (Martinet 1984).

Another definitional feature of the semiotic system of language is that chains of signs, known as *syntagmatic relations* in structuralism, form systematic, compositional patterns, allowing sentences to represent an open number of different events, with specific temporal and spatial features, i.e. *grammar* (Hockett 1960; Coseriu 1985; Trask 2004). Further, in order to create *texts* that form unified wholes, all languages have *cohesion* devices that help link clauses together to form, among other types of texts, coherent narratives, as described further in Section 2.3.

### 2.2.2. Gesture

Gestures can be defined as “expressive movements performed by the hands, the head, or any other part of the body, and perceived visually” (Zlatev 2015: 458). They can be performed either independently from language, or together with it as co-speech gestures: spontaneous movements of hands, and occasionally head and torso, that are performed while speaking (Goodwin 2003; Green 2014). To separate gestures from other bodily actions, on the one hand, and signed languages on the other, Andrén (2010) formulates a “lower limit” and “upper limit” of gesture (see Fig. 3). This division separates volitional interactional bodily expressions that constitute the semiotic system of gesture from bodily movements that lack “some required degree of *volition* (such as blushing) or some required degree or kind of *semiotic complexity* (such as taking an object offered by another person)” (Andrén 2010: 13). Signed language, on the other hand, is placed above the upper limit, because it is one specific variety of the semiotic system of language, with large vocabularies and complex grammars (Trask 2004).

Structurally, gestures have been analysed in terms of *gesture phrases*, which can further be divided into the following phases: *preparation*, *stroke*, *post-stroke hold*, and *recovery* (Kendon 1980, 2004; Kita, Van Gijn, Van der Hulst 1998). The stroke is what carries the gestural meaning in the gesture phrase, and each phrase can contain only one stroke. Other phases in a gesture phrase may or may not be present.

There are many functions that gestures serve in a polysemiotic utterance: regulating the interaction, structuring the discourse and expressing representational content (Kendon 2004). With respect to the latter, many gesture scholars make a threefold division between (a) iconic, (b) deictic, and (c) emblematic gestures, corresponding to the three basic kinds of semiotic grounds discussed in Section 2.1. This leaves out, however, gestures that “mark out, punctuate or some other way make reference to aspects of the structure of the discourse, either in

respect to its phrasal organization or in respect to its logical structure”, known as *pragmatic gestures* (Kendon 2004: 103).

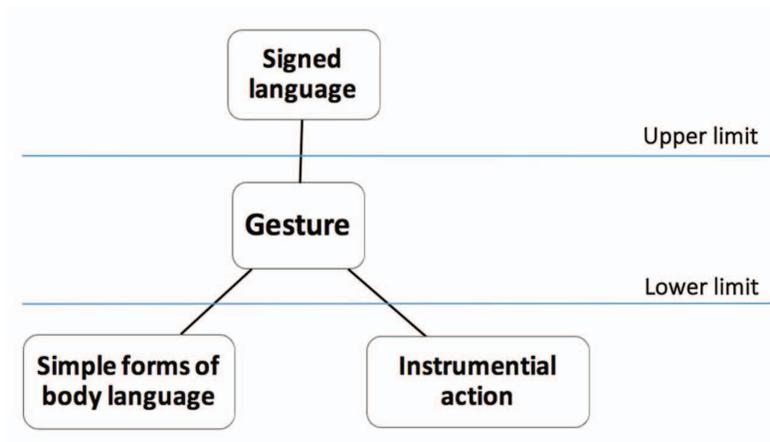


Figure 3. The upper and lower limits of gesture (adapted from Andrén 2010: 13).

*Iconic gestures* “where there is resemblance between the movements of the whole body, or parts of it, and properties of intended actions, objects or whole events” (Zlatev 2015: 461), have been extensively investigated, in particular in developmental and evolutionary contexts (Tomasello 2008). They can be divided into *enactments* and *representing gestures*, where in the former the body of the gesturer matches the action that is represented, and in the latter it does not, as when moving fingers do not represent fingers that move, but are taken to represent, for example, moving legs. The former can be expected to have a higher degree of primary iconicity. McNeill (1992) assigns iconic gestures a *viewpoint* distinction that the speakers/gesturers use to produce the movements: *character viewpoint (CVPT)* or *observer viewpoint (OVPT)*. In the former, the speaker enacts as though being “inside” the character; whereas the latter takes an external perspective. A mixture of the two viewpoints is also possible, for example, “when the hand is representing the trajectory of an inanimate entity, such as a ball that is flying toward a person, whereas the body of the speaker represents the person who is being approached by the ball” (Debreslioska *et al.* 2013: 450). While not identical, this division corresponds to that between enactments and representing gestures, described above.

*Deictic gestures* are those that “indicate or individuate an external target for an addressee, and include not only different types of pointing, but also acts which bring an object to the attention of the addressee (showing, giving, requesting)” (Zlatev 2015: 461). Abstract pointing includes the use of gesture to refer to an imaginary object instead of something in the vicinity (Andrén 2010).

*Emblematic gestures*, or emblems, are conventional in both form and meaning, and thus clearly qualify as symbols (Efron 1972[1941]; Andrén 2010). They can be defined as “those nonverbal acts which have a direct verbal translation, or dictionary definition, usually consisting of a word or two, or perhaps a phrase” (Ekman, Friesen 1969: 63). Just like linguistic signs, emblems differ across cultures in number, and form-meaning associations (Kita 2009). Examples of emblems shared in many Western cultures are, for example, the PEACE sign (formed with index and middle finger), the OK sign (circle formed with thumb and index finger), or GOODBYE wave.

As for *pragmatic gestures*, they have different functions such as (a) *performative* function – indicating whether the speech act is to be understood as a request, offer, rejection or something else; (b) *modal* function – indicating how what is said ought to be interpreted (e.g. expressing negation); and (c) *parsing* function – punctuating or structuring the speech (Kendon 2004), which may be used rhythmically to help emphasize the key aspects of the narrative (McNeill 1992).

### 2.2.3. Depiction

We may define depiction as a superordinate semiotic system (like language) that could be realized in a number of different media, from drawings in the sand to oil paintings or digital images, which “all involve the production of lines and patches of contrasting colours on surfaces in such a way to create a ‘likeness’ to real or imaginary objects or events, for pleasurable or edifying purposes” (Zlatev 2019). The products of depiction may be referred to as *pictures* (Sonesson 2010), which are by definition highly iconic signs, usually perceived visually (see Table 1). The lines and patches of colour that pictures are composed of are meaningless until they find their place as part of the pictorial representation. This, however, is distinct from the “duality of patterning” of language, as there is nothing corresponding to phonemes in pictures.

Sonesson (2011) analyses the picture sign in terms of Husserl’s notion of *pictorial consciousness*, which he develops further as involving four levels: (a) *picture thing* (the physical picture); (b) *picture object* (what is seen “in” the picture thing);

(c) *picture subject* (what is seen, but now with its “proper”<sup>5</sup> colours or details); and (d) *picture referent* (the thing in the world that is photographed or drawn) (see Fig. 4). Thereby, any critique addressed to the poor condition of the picture would be a reference to the material quality of the picture thing, but if the critique is instead addressed to the failure of the picture to present resemblance of the person or thing it portrays, the issue would be with the picture object. The picture referent, then, can always be found in the perceptual world if the referent indeed exists. Fictional characters in a picture book, however, would only have picture objects and subjects, but not referents.

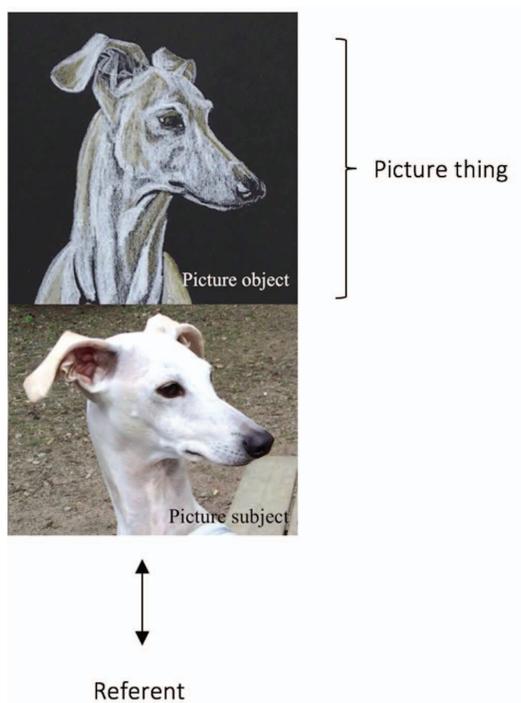


Figure 4. Pictorial consciousness as manifested in relation of *picture thing* (the art work), *picture object* (what is displayed in the drawing – a dog), *picture subject* (the dog in his real appearance), and the referent (Italian Greyhound named Chili, who exists in the perceptual world) (conforming the model of Sonesson 2010: 48).

<sup>5</sup> Even photographs that virtually have “an expression plane which is tautologously related to its content plane” are never fully able to reproduce reality in its true colours and light conditions, especially with the technology that enables enhancing the desired details and fading the undesired (Sonesson 2015b: 432).

Although a picture is clearly a type of iconic sign, it can also have the grounds of indexicality and conventionality.<sup>6</sup> The indexical relation between the photograph and its referents in the perceptual world at the moment the image was created is clearly an example of contiguity. The spatio-temporal indexicality between pictures in a sequence can also be used to represent causality: a picture of meal ingredients followed by a picture of a cooked dinner, or an image of a child kicking a ball and another of a broken window.

### 2.2.4 Polysemiotic communication

We discussed the semiotic systems of language, gesture, and depiction so far separately in order to highlight their different semiotic features and the nature of the medium they involve. However, most spontaneous human communication involves the combination of speech (or in the case of deaf people, signed language) and gesture, and in many cases also depiction, e.g. drawing. The combination of these semiotic systems in polysemiotic communication is spontaneous and allows complex interactions of sign use, where the different expressive potentials of the systems interplay with and balance one another in ways that remain to be explored in detail, especially in the case of polysemiotic narratives.

Some research in this direction uses the label of ‘multimodality’ (e.g. Jewitt 2014), but the term is used in different traditions with considerable ambiguity. In cognitive linguistics the term ‘modality’ is often used to refer more or less to what we mean by semiotic system, for example by calling verbo-pictorial images ‘multimodal metaphors’ (e.g. Forceville 2017). In gesture studies, speech and gestures are commonly referred to as ‘communicative modalities’, and language itself is often considered ‘multimodal’ (Vigliocco, Perniss, Vinson 2014). In the tradition of social semiotics (Kress 2010; Stöckl 2004), one considers the combination of ‘modes’ such as speech, text, picture, colour, music, typography, design etc. under the notion of ‘multimodality’. But, as pointed out by Green (2014: 10), this “leads to an abundance of modes that are difficult to compare”. Finally, ‘modality’ is often used to refer to the different senses: vision, hearing, touch and smell (and possibly others like proprioception), and perception is known to be multimodal.

Given this extensive ambiguity of the term ‘multimodality’, we constrain the use of the term to the synergy of two or more different sensory modalities in the act of perception (see also Stampoulidis, Bolognesi, Zlatev 2019). The significance

---

<sup>6</sup> Conventionality clearly manifests itself in different pictorial genres, but also in the equipment and styles used in the production of those images – nuances and contrasts (Sonesson 2015b).

of this kind of polysemiotic and multimodal orchestration becomes evident in face-to-face interaction, but is even more general. For example, it can be seen in the design of textbooks, websites and road signs, which may have (moving) pictures and sound in addition to written language. Until recently, the focus of communication research has predominantly been system specific, traditionally concentrating on written or spoken language. However, the importance of research that takes into account the interplay of different semiotic systems utilizing one or more sensory modalities has increasingly been acknowledged, as focusing on one system in isolation is not sufficient in order to understand the richness of human communication (Green 2014).

As is evident from the previous discussion, the semiotic system of gesture is tightly linked to the system of language when manifested as speech. Still, we consider gesture and language as closely interlinked but distinct semiotic systems. In this synergetic relationship gestures and speech interact closely with each other to create a richer form of *utterance* through a combination of two qualitatively different representations: spatio-motoric and linguistic (Kita, Özyürek 2003). Yet, the organization of the two modes of expression is flexible in a way that co-speech gestures can be employed differently depending on circumstances (Kendon 2004). Therefore, gestures are not to be regarded as “automatic” and as necessarily linked with speech, as argued by McNeill (1992), but as manifesting “deliberate expressiveness” (Kendon 2004: 15), even if the speaker/gesturer is not fully conscious of this.

### 2.3. Narratives and stories

The field of *narratology* (Todorov 1969; Genette 1980; Prince 2008) is exclusively devoted to the study of stories and narratives, but narratives are a topic of study within a variety of other disciplines as well. Not only are narratives a way to organize spatiotemporal events in a chain of causal episodes having a beginning, middle and end, but they also contain evaluations of the nature of these events and indicate the reasoning behind these evaluations (Labov, Waletzky 1967; Labov 1972; Kirstinä 2000). In this subsection, we explain relevant concepts, compare pictorial and verbal narratives, and introduce general structures and devices used in narrative construction. We limit ourselves to aspects that are used in the empirical study described in the rest of this article.

The ways the concept of narrative is defined in the literature are many, but a minimal definition is that “an object is a narrative if it is taken to be the logically consistent representation of at least two asynchronous events that do not presuppose or imply each other” (Prince 2008: 19). Further, such events need to

be temporally and causally connected. Following the classic distinction between ‘*fabula*’ (the story in its chronological order of events) and ‘*sujet*’ (narrations that may change this underlying order), both Berman and Slobin (1994) and Clark (2004) distinguish between ‘story’ and ‘narrative’ (e.g. Genette 1980; Bacon 2000).

An important dimension of narrative structure concerns the various components or aspects of a story or narrative. A common analysis involves breaking it up in *setting*, *theme*, *plot* and *resolution* (Sanford, Emmott 2012). The *setting* can include, for instance, the characters of the storyworld, the location(s) where the story events happen and also the passing or progression of time. The *theme* of a narrative expresses what the story essentially is about; it can be in the form of a goal to be achieved by the characters, or issues that the story raises, like death, spiritual growth or loneliness. The *plot* is an important aspect of the story, the red thread that is carried through the narrative. It is the sequence of events that leads up to the *resolution* – that is, the outcome. The plot structure is divided in three core components<sup>7</sup> in the study of Berman and Slobin (1994: 46): *onset*, *unfolding* and *resolution*. That is, for a plot to be meaningful, it needs a beginning (the onset), an end (the resolution) and events in the middle that lead up to the outcome (the unfolding).

Narrative research has predominantly focused on language (in literature and oral discourse) or, more recently, film and television, where events are represented in a time sequence (Ranta 2011). The importance of oral narratives is undeniably immense in human cultures, but other semiotic systems that are commonly used in modern societies to narrate are gestures and pictures, including picture sequences (Kress 2010). In the study described in this paper, pictorial narratives are compared with verbal narratives, and we may expect them to differ due to the properties of the corresponding semiotic systems.

The semiotic system of language is characterized by the sequential arrangement of elements in time, which provides the semiotic logic in verbal narratives. This unfolding of narrative information involves “linguistic *cohesion* on the micro level of individual clauses and adjacent clauses, and thematic *coherence* on the macro level of plot organisation” (Berman, Slobin 1994: 40, emphasis added). It is crucial to know who does what to whom when and where, which can be achieved by means of reference tracking. Languages offer several different mechanisms to indicate the information status of referred entities depending on their accessibility (Hickmann, Hendricks 1999). Another way to achieve cohesion in a verbal narrative is with help of connective devices, which can be used to link discourse

<sup>7</sup> As usual in the literature, the terminology is not fully calibrated. Berman and Slobin (1994) count resolution as part of the plot structure, whereas Sanford and Emmott (2012) consider the events that lead up to the resolution as plot.

segments together. Halliday and Hasan (1976: 244–270) provide a scheme of conjunctive relations in five categories depending on their function in the clause:

- *additive*: reflecting additive relation (e.g. ‘and’, ‘or’, ‘nor’);
- *adversative*: reflecting opposing relation (e.g. ‘although’, ‘but’, ‘yet’);
- *causal*: reflecting causal relation (e.g. ‘so’, ‘because’, ‘for this reason’);
- *temporal*: reflecting temporal/successive relation (e.g. ‘then’, ‘next’, ‘meanwhile’, ‘first’);
- *continuatives*: a rather diverse category including, for example, ‘well’, ‘anyway’ and ‘of course’, which bring cohesion in communication in different ways by indicating a response to expressed events, as in (1), reflecting attitudes or rhetoric, or dismissal of the previous utterance, as in (2).

(1) *They went to see if it could be behind the rock. **Well** it wasn't there either.*

(2) *Perhaps they were sleeping. **Anyway** the window was open.*

Clark (2004: 462) points out that “[n]arrators don’t construct a narrative *simply* to fit a story, selecting their words, phrases, and rhetorical devices to express the elements of a pre-determined conceptualization of events”. Rather, the content of a verbal narrative is limited to the specific aspects that the narrator *chooses* to describe. Thus, while bringing some pieces of information to the attention of the listener, any verbal narrative unavoidably leaves something out. There are several devices that can be used in speech in order to bring desired aspects of the story in focus while leaving less important aspects in the background. *Prosody* is one example of an effective foregrounding strategy (i.e. means to bring something in focus) in spoken narratives (Kress 2010). Sound is a rich element that offers a possibility for a considerable variation in energy, which can be used to stress elements of choice, accentuate words, or produce rhythm to organize speech. Variation in pitch produces intonation that is used, for example, to indicate attitudes or mark different kinds of phrases (i.e. questions or statements etc.), but also to *frame* – that is, mark if produced information units are new or given, and to tie these together into larger-level coherent units, a frame. Framing is essential for meaning-making in narratives, and different semiotic systems (and their subsystems) have their own ways of doing so. Other means that can be used for foregrounding (and backgrounding) elements in a spoken narrative are repetition, speed of narration, and the order in which information is presented, but also, by explicitly referring to events that are not happening, and the qualities that are not present in the storyworld may emphasize some aspects in the story that are (Prince 2008).

Fludernik (1996) argues that what is significant in any verbal narrative is the presence of a *narrator*, which typically reflects the consciousness/experience of the protagonist. Different narrative genres have their structural and stylistic conventions, and Fludernik distinguishes different types of oral narratives according to their level of experientiality. Two types that stand close to the opposite poles of the continuum of experientiality are *narrative report* and *narratives of personal/vicarious experience*. The former is a summary of events and action sequences with its main function simply to provide information. Reports tend to be objective descriptions of action and event sequences followed by the resolution, effectively offering “the ‘point’ of the story” (Fludernik 1996: 71). On the other hand, narratives of personal or vicarious experience are associated with “experiential value” either from a first or third person perspective (Fludernik 1996: 75–76).

As mentioned, pictorial narratives have received less attention in narrative research than verbal narratives (Sonesson 1997; Ranta 2011). Since there are no words in pictures to describe events and name characters, other means are used to present events. The semiotic system of depiction, as pointed out in Section 2.2.3, is predominantly grounded in iconicity rather than conventionality. While language uses the logic of time as a resource, still pictures use that of space. Pictorial narratives show elements that may not be present in a narrative presented by linguistic means, such as the location or size of objects, and the overall space. Thus, framed space is one of the affordances of pictures, as all elements are present simultaneously and arranged in the space in a way that conveys meaning about the relations between entities. Consequently, *transitivity* – that is, the relationship between agent and patient (i.e. who does what to whom) – is often directly given or interpretable in pictures (Sonesson 1997). Furthermore, while narrative is necessarily carried linearly in language, pictures afford much more freedom in the sense how they can be “read” (Seppä 2012). *Colour* is a powerful resource for framing in pictures, and can be used to highlight specific elements in pictorial narratives in a similar manner than some linguistic techniques can be used to highlight elements in speech (Kress 2010). Colour can thus be used to foreground elements in a narrative – bringing elements in focus.

The ability of static pictures to represent whole events and “tell stories” can be understood as matter of *secondary narrativity* (Stampoulidis 2019), where a single picture may “trigger” a pre-existing story/fabula. Two other types of pictorial narratives that differ in regards to narrative and temporal aspects are *continuous narratives*, and *narrative series* (Ranta 2011). Continuous narratives were represented already in historical works of art, where a single painting may show several different persons and events in the same pictorial space. In this type

of narratives different phases of an event series are represented simultaneously on a single canvas. Narrative series are most clearly represented in strip cartoons, where static pictures depicting different scenes or events are linked to form a narrative (Ranta 2011). These distinct pictures have a fixed reading order – often horizontal or vertical – and temporal continuum (indexical relationship) thus is more directly represented and linear (Cohn 2013).

The picture book *Frog, Where Are You?* by Mercer Mayer (1969) is a narrative series type of a pictorial narrative. It is not a strip cartoon, but a wordless book in which pictures represent events and together form a narrative that has a beginning and an end that are connected by events in between. When skimming the book, the first picture lets the viewer imagine a storyworld set in a bedroom in the evening with a boy, a dog, and a frog that is kept in a jar. When flipping the page to the second picture, the viewer does not create a new storyworld, but rather, adds a new scene to the previous one – a scene where the setting is otherwise the same but later at night (the boy and the dog are asleep). The successive pictures show changing settings of the *same* storyworld where the viewer is expected to add attributes and motivations that are not directly displayed in the pictures (Clark 2004).

Finally, as noted in Section 2.2, face-to-face communication is polysemiotic when the different semiotic systems act together in a way that, for example, speech names and gesture points or enacts. The interplay of the two semiotic systems acting together is also evident when it comes to managing coherence relations in oral narratives. It has been demonstrated that reference tracking is a bimodal and polysemiotic phenomenon (Levy, McNeill 1992; Gullberg 2006; Debreslioska 2019). Gesture space can be used effectively in order to create cohesion in a narrative. It can be done by “continued or recurring gestural patterns – handedness, hand configuration, or specific spatial area – associated with consistent visuospatial imagery or referential content over a stretch of discourse” (Gullberg 2006: 158). Hence, oral narratives are typically polysemiotic, combining properties of the systems of language and gesture. Gestures appear to facilitate the representation of location and space while speech is more effective in representing categorical relations among entities (Wagner Cook 2014). In effect, co-speech gestures may decrease cognitive load by shifting it from verbal working memory to other cognitive systems or external representations (Goldin-Meadow *et al.* 2001; Wagner, Nusbaum, Goldin-Meadow 2004).

## 2.4. Translating across semiotic systems

The term ‘translation’ is most commonly used to talk about translation from one language to another, but when meaning is transported from one semiotic system to another, for example, from pictures to linguistic form or from books to film, this involves the notion of ‘intersemiotic translation’ (Jakobson 1959), also known as ‘transduction’ (Kress 2010), or ‘intersemiotic transposition’ (Sonesson 2014).

Translation, both within and between semiotic systems is a much more complicated process than merely substituting one set of expressions for that of another, as already pointed out by Jakobson (1959). Sonesson (2014) characterizes translation as “a double act of communication” where the translator is “a doubly acting subject, as interpreter and as creator of a new text” (Sonesson 2014: 263–264). That is, the translator first interprets the message, and then makes a choice of either adapting to the creator of the message, to the audience, or else compromises by maintaining certain aspects of the original work, and adjusting others to fit better in the receiver’s background. Even in regard to translation from one language to another, the meaning can never be identically transferred in the process, but due to system specific properties, intersemiotic translation faces even greater challenges.

Each semiotic system has, as pointed out, typical semiotic characteristics and sensory modalities making intersemiotic translation a highly demanding and creative process. Sonesson (2014) argues that the process of transferring meaning from the semiotic system of language into that of pictures faces more challenges than the other way around, since language only has a limited set of qualities at disposal to be “abstracted from the wholes” of perceptual reality (Sonesson 2014: 274). That is not to say that the pictorial system would be able to deliver everything from perceptual experience, but that it reaches much closer to the perceptual world than language is ever capable of. Therefore, when translating a story from language to pictures, there may not be enough information in the source story available for the adequate transference of meaning, and many new details may need to be added, such as the way protagonists in the story look, the distance between them, the size and colour of entities, etc. Translating in the opposite direction faces other challenges, for it requires decisions on how to organize the narrative in terms of thematic hierarchies, while one also has to find the right words to describe the specific aspects of the meaning in the picture (Kress 2010; Sonesson 2014). Many elements represented iconically would be lost. Thus, when translating from pictures into language, co-speech gestures could be expected to help fill the gap.

Rimé and Schiaratura (1991) report evidence that iconic gestures are likely to be elicited when visual, motoric, or spatial information is translated into speech. We can readily make sense of this finding in terms of our framework: as gestures have a degree of iconicity that is intermediate compared to the semiotic systems of language (even in the spoken mode) and pictures, when translating from the latter to former, they can to some degree “fill in” for what language lacks.

## 2.5. Theory summary and hypotheses

This long section presented our theoretical framework and described human communication as inherently polysemiotic (in terms of semiotic systems) and multimodal (in terms of sensory modalities) in character. The concepts of narrative and story were defined, along with a number of narrative structures, in relation to narratives in different semiotic systems, and combinations of these. Finally, the notion of (intersemiotic) translation was taken up to present some of the many challenges that need to be met when stories are translated from one semiotic system to another.

As we have seen, we may have the same story/fabula, but when it is narrated using two different semiotic systems – speech and a sequence of pictures – we necessarily have two quite different narratives. The first question concerns how the semiotic system of a source narrative may influence the target narrative. The second question arises when such pictorial and verbal narratives need to be translated into polysemiotic narratives, using speech and gestures: would translation from speech lead to more *coherent* narratives, and would translation from a pictorial narrative lead to more perceptually detailed narratives, as may be expected given the affordances of the source systems and their sensory modalities? Thus, the following general hypotheses can be formulated:

- (1) Polysemiotic narratives produced as translations from a *verbal narrative* will have more narrative coherence, reflected in aspects such as a higher number of plot elements and more connective devices.
- (2) Polysemiotic narratives produced as translations from a *pictorial narrative* will have more iconic representations in gestures, and above all, enacting gestures.

## 3 Method

### 3.1. Design

We designed an experiment in which participants, divided in two groups, were first presented the same story as a pictorial (P) or a verbal (V) narrative, and then asked them to re-narrate the story to an interlocutor. Thus, the polysemiotic narratives elicited in this experiment are based on the same story, although they were presented in two different semiotic systems and sensory modalities.

#### 3.1.1. Materials

The pictorial narratives consisted of a sequence of 24 pictures, each representing one or more events. The verbal narratives consisted of an oral version of the same story, Mercer Mayer's picture book *Frog, Where are You?* (1969). For our study, colour was added to the original black-and-white pictorial narrative. Thus, the main characters and events became more easily foregrounded, in a similar vein that the semiotic system of speech affords foregrounding with the act of naming and prosodic variation (see Section 2.4).

For the verbal narrative, a script in English was downloaded from SALT Software LLC<sup>8</sup> (2015) webpage, and translated into Finnish (see Appendix B). This was then recorded as an audio file by a professional voice actor, making sure it contained expressive prosody in a way that fits the genre of storytelling.

#### 3.1.2. Participants and ethical considerations

Thirty-eight native Finnish speakers (25 women; 13 men), aged 20 to 53 years (mean 34) took part in the experiment in return for a cinema ticket. They were recruited via social media and personal acquaintances. The participants were divided evenly in groups based on the semiotic system of the narrative. Level of education (longer than 12 years), and gender (six men in the pictorial condition; seven men in the verbal condition) were controlled for in this grouping.

The participants signed a form of informed consent before the experiment, which included permission to be video-recorded. It was made clear that no personal narratives would be collected in the study, that the participants had the right to withdraw at any time, and that they would remain anonymous. The focus of the study was revealed at end of the experiment after the participant had filled in the feedback form (see Appendix C). The participants were also given a chance

<sup>8</sup> Frog, *Where are You?: English transcript*. Retrieved from: [http://www.saltsoftware.com/coursefiles/1202/FWAY\\_English.pdf](http://www.saltsoftware.com/coursefiles/1202/FWAY_English.pdf).

to discuss freely their experience and ask questions, which sometimes gave insight into their performance. These discussions, however, were not recorded.

### 3.1.3. Setup and procedure

The experiments were conducted in Finland, in the cities of Helsinki and Oulu, between February and April 2016. For the sake of ecological validity, naturalness and interaction, a number of factors were taken into account. Firstly, the location was selected so that possible outside disturbances could be kept minimal (e.g. a library or a home environment). Secondly, the aim was to make the discourse feeling as natural as possible between the addressee (i.e. the first author or a research assistant) and the narrator. Therefore, the participant was an acquaintance (i.e. a friend, a classmate or a neighbour) of either the first author or the research assistant.

Participants were told that the study compared understanding a story presented in two different ways – in speech or in pictures. They were advised to attend to the respective narrative carefully, because they would have to retell it to the addressee in their own words right after the exposure. Each participant was exposed to the narrative only once. The length of the audio file was three minutes and 30 seconds, and in order to make the length of the narratives identical between the groups, the sequence of pictures was shown in the same pace as the audio recording. There was only one picture per folding. The method of the controlled page turning differs from the “normal” narrative elicitation procedures based on the Frog story, in which the participants can use the pictures as props for their narratives by freely turning the pages back and forth in their own tempo when narrating. This alternative method was necessary in order to make the narratives in the two systems comparable.

Once the first part of the experiment – that is, hearing or seeing the story – was over, the participants were asked to proceed to reproducing, in effect *translating*, the narrative to the addressee. After the elicited narrative task, the participants were asked to fill in a feedback form, to provide some background information concerning gender, age, handedness, language skills and possible time lived abroad; and feedback concerning how the experiment had been experienced – that is, whether the task had been challenging, and whether it could have been easier or more difficult had the participant experienced the story in the alternative semiotic system.

## 3.2. Speech annotation and coding

The reproduced narratives were video-recorded, and later transcribed and annotated in detail using the multimedia annotator ELAN 4.9.1-b, a professional

software for the creation of multiple layer annotations on video and audio resources (Wittenburg *et al.* 2006). When repetition occurred as a result of disfluency, the resulting speech (or speech-gesture combination) was coded only once. Furthermore, incomplete words, interruptions and “pause fillers” (e.g. ‘mmm’, ‘hmm’, ‘ööö’) and false starts were deleted.

### 3.2.1. Clauses

The basic unit of analysis was the clause, defined by Berman and Slobin (1994: 657) as “any unit that contains a unified predicate”, that is, “a predicate that expresses a single situation (activity, event, or state)”. Clauses classified this way consist of only one verbal element<sup>9</sup>, as in (3), or two as in (4). Likewise, embedded clauses were treated as separate clauses, as in (5).

- (3) *hyppäsi ulos ikkunasta*  
‘jumped out of the window’
- (4) *poika pyysi koiraa | olemaan ihan hiljaa*  
‘he asked the dog | to be quiet’
- (5) *ja ne meni { sen puun jossa se mehiläispesä oli kiinni } sen alle*  
‘and they went { the tree where the beehive was attached to } under it’

The clauses were divided into *narrative*, *metanarrative* and *extranarrative* units. What is classified as a unit of the narrative kind contains information that is part of the story (Mushin 2001), even when this was inferred rather than directly given in the source, as in (6) and (7).

- (6) *ja sit se sammakko karkas*  
‘and then the frog escaped’
- (7) *no siitähän ne mehiläiset suuttu tietenkin*  
‘of course the bees went mad because of that’

A unit of the metanarrative kind represents a personal stance (e.g. evaluations of the perceived narrative), as in (8) and (9).

<sup>9</sup> In the example given by Berman and Slobin (1994: 657): “[...] clauses will be comprised of a single verbal element; however, infinitives and participles which function as complements of modal or aspectual verbs are included with the matrix verb as single clauses, e.g. *want to go, started walking*”.

- (8) *mutta mitä ihmettä sitten sen jälkeen tapahtu?*  
 ‘but then what happened after that?’
- (9) *mut sitä mä en ihan ymmärtäny*  
 ‘but that I didn’t quite understand’

Extranarrative clauses, finally, make a comment upon the stimuli or the experiment situation, as in (10) and (11).

- (10) *sitä kuva ei paljastanu*  
 ‘the picture didn’t reveal that’
- (11) *siitä puuttu sellanen kohta missä...*  
 ‘it was missing an episode where...’

### 3.2.2. Plot elements

Narrative clauses were divided in different plot components according to the coding scheme provided by Berman and Slobin (1994), introduced in Section 2.4.2:

- (1) *the onset* (the boy realizing the disappearance of the frog);  
 (2) *unfolding* (search for the missing frog);  
 (3) *the resolution* (the boy finds the frog).

This coding scheme was modified to take into account the whole story from introduction to ending (which were left out in the original coding scheme). Since introduction of the topic and ending of the story are important parts of the narrative structure in verbal narratives, conventional storytelling expressions such as ‘once upon a time there was’ were kept and coded as narrative units in the modified scheme. Following this and the clause unit distinctions described in the previous paragraph, the clauses were divided into *introduction*, *onset of the plot*, *unfolding of the plot*, *resolution of the plot*, *ending*, and in *metanarrative*, and *extranarrative* units. This division allowed the narrative and non-narrative units to be analysed separately.

### 3.2.3. Connecting devices

Clause initial connective devices were identified and counted in each group following the scheme provided by Halliday and Hasan (see Section 2.4). The connective devices were analysed individually, because often they occurred as strings of conjunctions, and they were categorized based on the function of the

particular combination. It was found that the function of a connective device in a clause was not always stable and depended on the context. As can be seen in Table 2 (Appendix A), the conjunction *että* ‘that’<sup>10</sup> appears in two different categories.

### 3.3. Gestures

All strokes [meaningful part of the gestural movement where the excursion of the limb reaches the apex (Kendon 1980, 2004)] were identified, and distinguished as having predominantly an *iconic*, *deictic*, *emblematic* or *pragmatic* function. Iconic gestures were divided further in two categories: *enacting* or *representing* (see Section 2.2.2). Only gestures that were produced in narrative clauses were considered for analysis.

### 3.4. Specific hypotheses

Based on the potentials/characteristics of each system, we derived the following specific hypotheses. H1 and H2 are operationalizations of the first general hypothesis regarding narrative coherence, and the H3-H4 are operationalizations of the second general hypothesis that considered the use of iconic gestures (see Section 2.5):

- (H1) There will be a higher number of plot elements when translated from verbal than pictorial narratives due to linguistic cohesion in the source narrative, which provides the foundation for the (linear) unfolding of the narrative plot.
- (H2) There will be a more diverse use of connective devices (in terms of function) when translated from verbal compared to pictorial narrative, due to their presence in the speech source narrative, whereas in the picture source narrative the relations between successive events need to be inferred.
- (H3) There will be a higher number of iconic gestures when translating from pictorial than verbal narratives, reflecting the more iconic nature of the source narrative.
- (H4) There will be more enacting gestures when translating from pictorial compared to verbal narratives, as these are the kind that most closely correspond to the primary iconicity of pictures in the source narrative.

---

<sup>10</sup> *‘Että’*-conjunctions that introduced direct speech were placed in *continuatives* and those that were found to function as causal connectives were placed in the *causal* category.

## 4. Results

### 4.1. General results

The results are collected in Table 2, and statistical analyses in Table 3. The statistical results are based on hierarchical Poisson regression (for rates) and logistic regression (for proportions). The main predictor was always the semiotic system (i.e., depiction vs. language). In some of the analyses, clause length was included as a covariate. In all analyses, finally, participants were included as a random predictor.

Table 2. Overview of the results in pictorial (P) and verbal (V) conditions.

	Total		Proportions		Rate per clause (per word)	
	P	V	P	V	P	V
<b>Narratives</b>	<b>19</b>	<b>19</b>				
<i>Words</i>	7848	6822				
<i>Clauses</i>	1285	1117				
<i>Average clause length (words)</i>	6.10	6.10				
<b>Narrative clauses</b>	<b>1237</b>	<b>1042</b>	<b>0.97</b>	<b>0.93</b>		
<i>introduction</i>	136	61	0.11	0.06		
<i>onset of the plot</i>	86	72	0.07	0.07		
<i>unfolding of the plot</i>	895	763	0.72	0.73		
<i>resolution of the plot</i>	109	141	0.09	0.14		
<i>ending</i>	11	5	0.01	0.00		
<b>Metanarrative clauses</b>	<b>42</b>	<b>66</b>	<b>0.03</b>	<b>0.06</b>		
<b>Extranarrative clauses</b>	<b>6</b>	<b>9</b>	<b>0.00</b>	<b>0.01</b>		
<b>Connective devices</b>	<b>931</b>	<b>719</b>			<b>0.69</b>	<b>0.75</b>
<i>additive</i>	433	318	0.46	0.44		
<i>adversative</i>	42	75	0.05	0.10		
<i>causal</i>	84	46	0.09	0.06		
<i>temporal</i>	295	253	0.32	0.35		
<i>continuatives</i>	77	27	0.08	0.04		
<b>Gestures</b>	<b>321</b>	<b>534</b>	0.01	0.01	<b>0.26</b>	<b>0.51</b>
<i>deictic</i>	3	6	0.04	0.01		
<i>emblematic</i>	14	4	0.71	0.78		
<i>iconic</i>	229	418	0.20	0.02		
- <i>enacting gestures</i>	46	10	0.80	0.98		
- <i>non-enacting gestures</i>	183	408	0.23	0.20		
<i>pragmatic</i>	75	106				

Table 3. Statistical difference between pictorial (P) and verbal (V) conditions in H1-H4. The table shows the estimate of the effect (EST), its standard error (SE), the *z*-value and the *p*-value.

	EST	SE	<i>z</i>	<i>p</i>
H1 Plot elements	-0.799	0.467	-1.712	0.087
H2 Connective devices	-0.193	0.108	-1.782	0.075
<i>additive</i>	0.049	0.180	0.274	0.784
<i>adversative</i>	-0.873	0.243	-3.585	0.000
<i>causal</i>	0.286	0.239	1.197	0.231
<i>temporal</i>	-0.047	0.219	-0.216	0.829
<i>continuatives</i>				
H3 Iconic gestures	-0.199	0.366	-0.543	0.259
H4 Enacting gestures	2.303	0.599	3.846	0.000

#### 4.2. Plot elements (H1)

As can be seen in Table 3, the narratives derived from the pictorial condition were longer in comparison with those produced in the verbal condition. Although there were no differences in the average clause length between the conditions, the participants who had seen the story in pictures produced in total almost 200 narrative clauses more than the participants who had heard the story. Proportionally, the difference in the number of plot elements between the two conditions is marginally significant (see Table 4). However, the difference goes in the opposite direction of what we expected. Thus, H1 was not supported. Interestingly, there were more metanarrative and extranarrative clauses in the verbal source narrative condition.

#### 4.3. Connective devices (H2)

Overall there were higher rates of connective devices in the pictorial than in the verbal condition, a difference that is marginally significant (see Table 4). The most frequently used connective devices in both conditions were of the *additive* category, followed by the *temporal* category (see Fig. 5). Statistically, the use of connectives in those two categories and in the *causal* category did not differ significantly. However, the connective devices in the *adversative* category were used twice as often in the verbal than in the pictorial condition, whereas the opposite was true for the connectives in the category of *continuatives*. Thus, while H2 cannot be said to be supported, as the results did not show a more “diverse” use of connective devices in terms of function in the verbal compared to the

pictorial condition, the differences in adversatives and continuatives could be due to the source narrative. We return to this in Section 4.6.

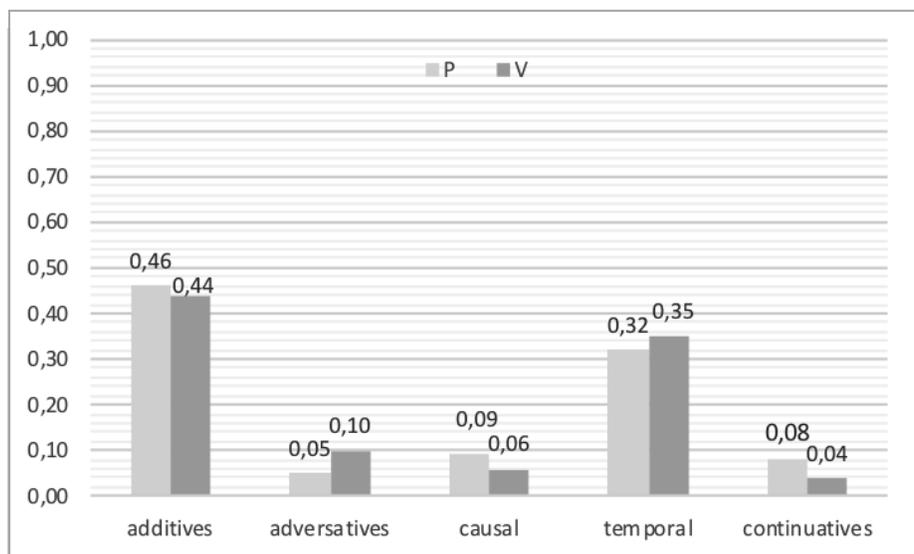


Figure 5. Proportions of connective devices in the five different categories of conjunctive relation between pictorial (P) and verbal (V) conditions.

#### 4.4. Iconic gestures (H3, H4)

As can be seen in Fig. 6, the frequency of gestures in the verbal condition was twice the frequency of gestures in the pictorial one. Iconic gestures were proportionally the largest group in both conditions when dividing gestures in different types according to their main function: deictic, emblematic, iconic, or pragmatic (see Fig. 7). Although there is a substantial difference in the frequency of iconic gestures between the two conditions, opposite to what we expected, (see Table 4), the difference is not statistically significant when the number of narrative clauses in each condition is taken into account (see Table 3). Although there were twice as many iconic gestures per clause in the verbal condition compared to the picture condition, the frequency of enacting gestures nevertheless was higher in the latter (see Tables 2 and 4), in which 20% of the iconic gestures were enactments, compared to only 2% of the iconic gestures in the verbal narrative condition.

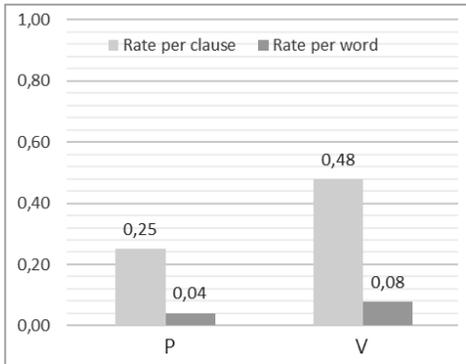


Figure 6. Gesture rate per clause and per word between pictorial (P) and verbal (V) conditions.

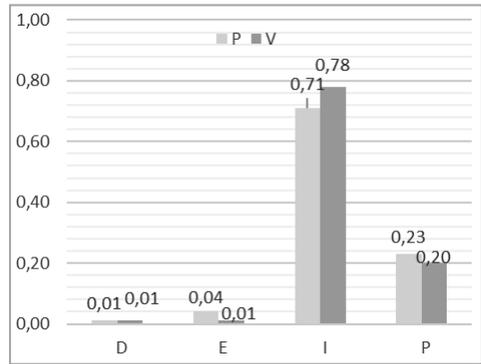


Figure 7. Proportional division of gestures in functions: deictic (D), emblematic (E), iconic (I), and pragmatic (P) between pictorial (P) and verbal (V) conditions

Table 4. Proportions of iconic gestures counted from the total number of iconic gestures across conditions (637 gestures); then in each condition when compared to the other gestures in the group (321 in pictorial (P) and 534 in verbal (V) condition); then per narrative clause, and per word.

	P	V
All iconic gestures	0.35	0.65
All gestures in the condition	0.71	0.78
Per clause	0.19	0.40
Per word	0.03	0.07

## 5. Discussion

Although three of the four hypotheses did not receive support, the results presented in the previous section nevertheless indicate that the semiotic system of the source narrative may lead to differences when translating into polysemiotic narratives.

The proportion of plot elements was not higher when translating from speech, compared to translating from pictures (H1). While the results did not show “more diverse” use of connectives (H2), the significantly different proportions of adversatives and continuatives in the two conditions were indicative of differences in the use of connective devices. The number of iconic gestures in the verbal

narrative condition was nearly twice the number of those in the pictorial narrative condition (contra H3). The only hypothesis that was clearly confirmed was H4, which predicted more enactment gestures when translating from the pictorial than the verbal narrative.

As noted, the participants who had experienced the story in pictures produced longer narratives than the participants who had experienced the story in speech. Although language may afford more freedom in terms of how one may imagine the perceptual storyworld (e.g. the appearance of the characters), in comparison to pictures where the perceptual details are directly given, pictorial narratives could still be argued to afford more freedom in another sense: they allow the narrator to improvise with the narrative details in a way allowing for a more creative story around certain core elements given in the pictures.

One clear difference in the plot elements between the conditions was found in the number of clauses produced for the introduction to the story. The participants who had heard the story tended to move directly to the onset of the plot after giving a brief introduction of the main characters and possibly of the setting, which often only involved a mention of the frog's jar, or that the boy and the dog were sleeping. On the other hand, the introductions produced by the participants who had seen the pictorial narrative usually contained more detailed descriptions of the setting including the spatial arrangement of the characters or details of the surroundings and the characters, and descriptions of the activities the characters were engaged in.

Such dissimilarities could be explained by the differences in the source semiotic systems. As was explained in Section 2, there are more perceptual details present in pictures than in words and sentences, and framed space is one of the affordances of the semiotic system of pictures that language lacks. Moreover, although words can induce different mental imagery, pictures can provide a no less rich, and possibly richer, setting for narrative hypothesising, giving rise to “theories” and questions. This is related to another factor that explains the higher rates of introductory clauses in the pictorial condition: some participants produced an introduction that preceded the first scene in the story, explaining why the story started from certain settings, as in (12):

- (12) *olipa kerran Ville ja sen koira Putte . he olivat metsässä sitten olleet . etsimässä erilaisia käpyjä . ja kuin ollakkaan niin oli löytynyt sitten sieltä tommosen puun kolosta niin sammakko. no Villehän päätti sit sen sammakon ottaa . ja pistää lasipurkkiin . ja viedä kotiin . ja tää Putte-koira oli ihan ihmeissään . et miten tää sammakko voi tääl purkissa olla.*

‘once upon a time there was Ville and his dog Putte . they had been in the woods . looking for different kinds of cones [strobili] . and how about that in the hollow of a tree they had found a frog. well Ville of course decided to take the frog . and put it in a glass jar . and take it home . and Putte the dog was in a state of utter bewilderment . [wondering] that how can the frog be in this jar ?’

The participants who had heard the story produced non-narrative clauses in order to indicate what was found to be incoherent in the source narrative, as in (13), or to point out flaws in their own production of the story, as in (14). This could explain the higher rates of non-plot elements in this condition.

- (13) *ja sit semmonen niinku kummallinen kohta siin sadussa oli siis vaan se . jos mä saan kommentoida . et siin ei ollu mitään semmosta . et poika riensi rappusia alas tai näin . mul meni vähän keskittyminen siinä kohtaa .*  
 ‘and then a weird part in that story was just . if I can comment . that there wasn’t anything like . that the boy rushed down the stairs or so . my concentration suffered a bit at that point.’
- (14) *mä kyl oon unohtanu jonkun vaiheen tässä .*  
 ‘I have forgotten some part here now .’

When evaluating the “fluidity” of the narrative performance, the narratives perceived visually seemed to have been internalized to a higher degree than those perceived by hearing. The participants in the verbal narrative condition more often seemed to be restricted by the source narrative system in their own retelling performance. This could be seen, for example, in uncertainty in the choice of words or in the order of the events. Such restrictions could lead to shorter narratives and explain further the higher rates of metanarrative clauses, such as (15).

- (15) *ja en muista miten se liittyy tähän tarinaan .*  
 ‘and I can’t remember how it’s related to this story’

The reason for this difference between the conditions could be that language is more descriptive and characterized by the sequential arrangement of elements in time, whereas pictures are more demonstrative and mimetic, and the information in a picture is available simultaneously (see Section 2.4.3). Thus, although the pacing of the scenes was controlled and equal between the conditions, the unfolding of information between them differed. This seemed to have an effect

on how the story was internalized in each condition. Although not statistically significant ( $EST = 0.025$ ,  $SE = 0.329$ ,  $t = 0.077$ ,  $p = 0.939$ ), the difference between the conditions was seen also in the narration speed: participants in the pictorial condition narrated slightly faster (0.05 min/clause) than the participants in the verbal condition (0.06 min/clause). The intersemiotic translations in the pictorial condition were longer in terms of words and clauses than those in the verbal condition, and yet in total the two conditions took roughly the same time to narrate (61.78 minutes in P vs. 61.30 minutes in V), which supports the view that the narratives resulting from the pictorial condition were more fluent. This was also confirmed by the participants' evaluation of their own performance in the feedback form given at the end of the experiment: 11 out of 19 participants in the verbal condition regarded the task as challenging, and 15 believed they would have remembered the story events better had they seen the story in pictures. In the pictorial condition only six participants out of 19 regarded the task "slightly" challenging, and only two believed they could have understood or re-narrated the story better if they had heard the story instead of seeing it.

In the verbal condition, uncertainty with word choices seemed to be based on attempts to remember the right words the voice actor had used in the source narrative, as in (16). In the pictorial condition, the participants also had difficulties with word choices, but these were rather based on identification of the animals represented in the pictures, as in (17). This particular example also shows how picture narratives may get different interpretations, because the participant here refers to the event represented in Figure 8 – usually interpreted as the gopher biting the boy's nose – as the boy's reaction to bad smell.

(16) *ja se koira kiinnitti huomion johonkin ampiaispesään . vai oliko se mehiläispesä ?*

'and the dog paid attention to some wasps' nest . or was it a beehive ?'

(17) *ja sitten oli joku myyrä tai tämmönen pikkuelukka . joka tuoksu pahalle . oisko se ollu joku haisunäätä tai jotain vastaavaa .*

'and then there was some vole or this kind of a small creature . that smelled bad . maybe it was a skunk or something .'



Figure 8. Picture 10 in the picture source narrative that presents the gopher biting the boy's nose while the dog is harassing the bees (adapted from Mayer 1969).

Consistent with the differences in the level of internalization of the story content, and the unfolding of information in the source narrative, it also makes sense that a narrator that narrates based on the speech source narrative might easier confuse who did what to whom, as in (18), while in the corresponding picture of the picture source narrative (see Fig. 9), the correct position of the boy is more likely to be remembered. Accuracy in story details was not among the tested factors, but it nevertheless played a role in the experiment, for, as demonstrated above, uncertainty often resulted in hesitation or manifestation of non-narrative clauses.

- (18) *joka [peura] sit niinku nappas jotenkin sarviensa väliin sen koiran . ja lähti juoksee . ja sit se poika lähti niinku juoksee sen peuran perässä .*  
'who [deer] then like grabbed the dog somehow between its antlers . and started running . and then the boy like started running after the deer .'



Figure 9. Picture 16 in the picture source narrative (adapted from Mayer 1969) that corresponds with line 16 in the English transcript (see Appendix B): “The deer started running with the boy still on his head. The dog ran along too. They were getting close to a cliff.”

The greater freedom of interpretation afforded by the semiotic system of pictures – compared to that of language – can result in different ways of telling the “same” story, as was explained in 2.4.3. This was also demonstrated in our study, for many of the narrators in the pictorial condition freely created a personalized story by naming the characters and giving various motivations for their actions, as several of the excerpts in this section illustrate.

Consistent with this, one of the most prevalent characteristics differentiating the two narrative conditions from one another was the level of *experientiality*. The noticeable aspect that could be linked to experientiality in the polysemiotic narratives is the presence of the consciousness of the protagonist or the consciousness of the narrator (see Section 2.4.3). Interestingly, this was found to be more typical in pictorial than in verbal condition, and was manifested in the frequency of direct speech, reflecting character consciousness (i.e. shifting the perspective from third person narration to first person narration). There were 74 instances of first-person narrative clauses in the picture condition and only 12 in the speech condition, with significantly different proportions (see Fig. 10).

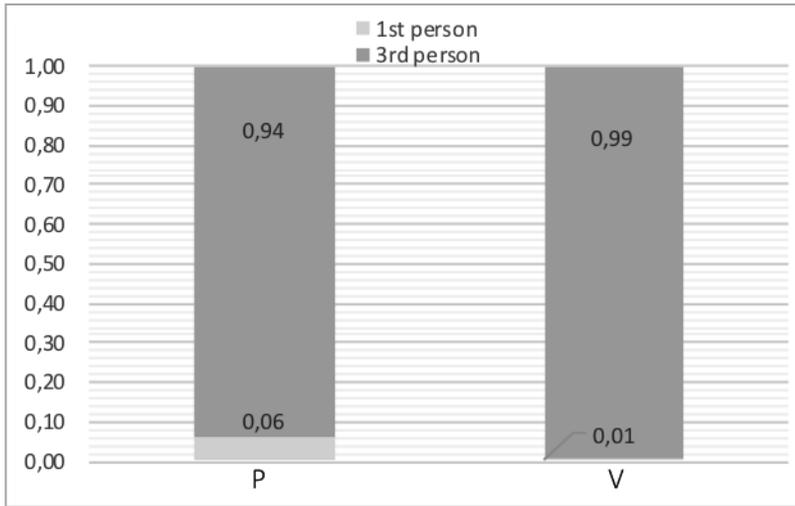


Figure 10. Proportion of first-person direct speech and third-person indirect speech perspectives. The difference between pictorial (P) and verbal (V) conditions is significant (EST = -1.573, SE = 0.495,  $z = -3.176$ ,  $p < 0.001$ ).

Mostly, the instances of direct speech in the verbal condition were those that were also present in the source narrative: the boy calling the frog, as in (19). In the pictorial condition the perspective was chosen more freely between any of the main or peripheral characters. Here, the narratives were also more often found to reflect “experiential value” through empathy, as in (20).

(19) *missä olet sammakko ?*  
 ‘where are you frog?’

(20) *ja eihän Max kun ei ollu aikasemmin nähny . ni ei tienny . et se oli ampiaispesä .*  
 ‘and of course since Max had never seen alike before . didn’t know . that it was a wasps’ nest.’

In line with this interpretation, *adversatives* were more frequent in verbal condition and *continuatives* in the pictorial one. Firstly, the most frequent adversative *mutta* ‘but’ was in many occasions found to be preceded by a metanarrative or extranarrative clause, as in (21).

(21) *sitä tarina ei kertonut . mutta poika oli onnellinen .*  
 ‘the story didn’t tell that . but the boy was happy.’

Further, the most frequent continuative conjunction *että* ‘that’ was often used to introduce direct speech, as in (22), and since the majority of first-person perspective narration happened in the pictorial condition, this could explain the difference in the continuatives between the conditions.

- (22) *että* älä mee sinne . *et* ampiaiset tulee . ja pistää sinua .  
 ‘that don’t go there . that the wasps will come . and sting you .’

Since the target narratives produced in the verbal condition were shorter in terms of narrative clauses, yet combined with a higher number of gestures and took more time to narrate than the narratives in the pictorial condition, a closer inspection of the gestures in connection to speech was required. As we have seen, the demand on memory and the ability to internalize the story has an effect on the level of apparent narrative experientiality.

In Section 2.3 it was suggested that gesturing may help shifting load from verbal working memory to other cognitive systems or external representations, and the findings of this study could indeed indicate a facilitative role of gestures – namely, that of iconic representing gestures – on narrative production. The listener of a spoken narrative does not have the freedom to construct a personalized story the way a viewer of a picture narrative has, thereupon she/he is required to rely on memory in a different manner in the retelling of the story. As mentioned, in terms of Fludernik’s (1996) continuum of experientiality, the narratives in the verbal condition tended to have more properties from the *narrative report* style being built upon “second-hand experience or on a summary of first-hand experience rendered non-experientially” (Fludernik 1996: 71). Interestingly, the gestures of the verbal condition participants were found to reflect this level of experientiality. The narrators in this condition more often than in the pictorial one narrated the story from the “outside” of the storyworld, and in a way by help of gestures drew a visual map of the story events and characters in front of them. This systematic use of gesture space could be seen as a way to create and maintain cohesion in the polysemiotic narrative, which is helpful not only for the interlocutor, but also for the narrator in the process of narrative production.

The narratives in the pictorial condition, in contrast, seemed to have more properties from *narratives of vicarious experience* in terms of Fludernik’s continuum, which could be seen, for example, in a greater use of the first-person perspective in both gestures and speech (thus reflecting a high level of experientiality). Instead of drawing a map of the events in front of them, the narrators seemed to position themselves *inside* the storyworld. This explains the occurrence of first-person enacting gestures (and direct speech) in the picture narrative condition.

Gaze was one of the indicators of the narrator's position either inside or outside the storyworld. This outside- *versus* inside-the-story perspective could be the reason for the substantially different rates of gestures between the two conditions (although the difference did not reach statistical significance). Namely, managing the visual map of the storyworld from the outside would require a greater use of gestures than when one is positioned inside the story, being able to refer to things in relation to one's own body. Correspondingly, narrating a story based on second-hand experience is more challenging than being able to express the events through access to the consciousness of the protagonists – that is, through perceptual focalizations, or through empathy.

As mentioned above, direct speech and narration through empathy reflecting the presence of a conscious actor (i.e. manifestation of a higher level of experientiality) was often combined with the first-person perspective in gestural expression, which is possibly the reason for the higher occurrence of iconic enacting gestures in the pictorial condition compared to the verbal one. This evidence links with Rimé and Schiaratura's (1991) finding that iconic gestures are likely to be produced when visual, motoric, or spatial information is translated into speech. However, in light of our findings, this can only clearly be said about iconic *enacting* gestures, not concerning iconic representing ones.

To summarize, only one hypothesis received clear support, but the findings nevertheless suggest that the semiotic system of the source narrative indeed appeared to have an effect on the target narratives. The semiotic system of language (specifically, its subsystem of speech) resulted in shorter polysemiotic narratives in terms of clauses, but a longer narration time and a higher rate of iconic representing gestures, possibly reflecting a lower level of experientiality. In effect, the perspective on the story events was more distant, leading the narrators to narrate the story from the outside of the storyworld, in which they got support from a systematic use of co-speech gestures. On the other hand, the semiotic system of depiction reflected iconicity in the retelling performance through presence of more iconic enacting gestures. The narratives in this condition were characterized by a first person perspective in both gestures and in speech, indicating a high degree of experientiality. The vividness of the narratives stemming from the pictorial condition is also consistent with the greater semiotic freedom afforded by pictures. The verbal condition instead resulted in more narratives in the style of report, consistent with the suggestion that the semiotic system of language allows less semiotic freedom of interpretation and arguably makes more demands on memory.

When translating from monosemiotic and unimodal narratives realized as either language or pictures into multimodal polysemiotic narratives, strictly

speaking, only the story in pictures is “moved” from one system to another, whereas the story in language only “adds” the semiotic system of gestures. As we have seen, gestures in these two conditions were used differently in their interaction with speech, which supports the view discussed in Section 2.2.4 that language and gesture are two distinct but closely interacting semiotic systems (instead of constituting a single system). That is, spatio-motoric representations combine with linguistic representations for a better narrative expression suitable for each situation.

## 6. Conclusions

This article centred around the *polysemiotic* and *multimodal* nature of human communication, which are conceptually and empirically distinct notions. We examined how narratives in different semiotic systems and sensory modalities are translated into polysemiotic narratives, and investigated the influence of the source semiotic system on this process. What happens when the same story expressed in either language or pictures is translated into a polysemiotic narrative? Since the two semiotic source systems have different potentials for expressing meaning, it was expected that some system-specific elements would transfer to the target narratives delivered in speech and co-speech gestures. Thus the two different “retellings” of the same story would lead to different polysemiotic narratives. This expectation was supported, and the source narratives translated into polysemiotic narratives were different, for example, in length, in the use of gestures, and in terms of creativity.

A second research question asked whether hearing the story would result in more coherent retellings in respect to the organization and the development of the plot. While the results do not indicate that the narratives translated from the speech narrative would result in more coherent narratives than those translated from the picture narrative, they do suggest some differences between the conditions in the use of devices that are employed to create cohesion. Furthermore, more side comments (in non-narrative clauses) were used in the speech narrative condition to support the development of the plot. Thus, it can be concluded that hearing a story does not necessarily result in a more coherent polysemiotic narrative compared to seeing the story in pictures.

A third research question asked if seeing the story would give rise to more perceptually detailed narratives, namely, iconic gestures. Contrary to expectations, more iconic gestures were in fact found in the narratives translated from the speech narrative, but these were nearly entirely non-enacting in type. However, iconic first-person, enacting gestures were indeed significantly more frequent in

the picture condition, and these enactments occurred often in connection with direct speech, which contributed to making the narratives of this group livelier.

The reason for the differences in the use of iconic gestures was apparently that the narrators in the speech condition had an outside-the-story perspective in their narration, which required more systematic use of gesture space in order to represent developments of the events and track referents, whereas the narrators in the picture condition used an inside-the-story perspective, and were able to refer to things in relation to their own body. This condition-specific difference in perspective could be linked to the differences in the level of experientiality, which arose as the most dominant feature differentiating the intersemiotic translations from one another. A number of consistent factors indicated that the level of experientiality was lower in the speech narrative condition, reflected in fewer cases of first-person perspective in both speech and gestural expression than in the narratives translated from the pictorial narrative. The polysemiotic narratives in the picture condition often resulted in more personalized improvisations in which the narrative “added” to the story. In many occasions the narratives translated from speech mostly provided information in a report style about the characters, the main events, and about the resolution, whereas in the picture narrative condition the retelling performance was more comparable to genuine storytelling. This difference can be traced back to the source narrative systems with language being stricter in terms of “accuracy”, whereas the pictorial system affords a greater semiotic freedom of interpretation. In general, the story was not as well internalized in the speech condition, which was reflected in many cases of hesitation and uncertainty that affected how the story was narrated. On the whole, the results indicate that different source semiotic systems may indeed affect how polysemiotic narratives are constructed.<sup>11</sup>

---

<sup>11</sup> **Acknowledgments.** We would like to thank the participants in the study, the confederates, and the Finnish voice actor Anna Kiuttu, who read the Frog Story, for their help and valuable time. We also thank Georgios Stampoulidis, Simon Devylder and Göran Sonesson for providing feedback on earlier versions of this paper. In addition, we thank an anonymous reviewer and the editors of the journal for a number of helpful comments and suggestions.

## References

- Ahlner, Felix; Zlatev, Jordan 2010. Cross-modal iconicity: A cognitive semiotic approach to sound symbolism. *Sign Systems Studies* 38(1/4): 298–348.
- Andrén, Mats 2010. *Children's Gestures from 18 to 30 Months*. Ph.D. dissertation. Lund: Centre for Languages and Literature, Lund University.
- Bacon, Henry 2000. *Audiovisuaalisen kerronnan teoria*. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Berman, Ruth; Slobin, Dan I. 1994. *Relating Events in Narrative: A Crosslinguistic Developmental Study*. New Jersey: Lawrence Erlbaum Associates.
- Clark, Herbert H. 2004. Variations on a ranarian theme. In: Strömquist, Sven; Verhoeven, Ludo (eds.), *Relating Events in Narrative: Typological and Contextual Perspectives*. London: Lawrence Erlbaum Associates, Publishers, 457–476.
- Cohn, Neil 2013. Visual narrative structure. *Cognitive Science* 34: 413–452.
- Coseriu, Eugenio 1985. Linguistic competence: What is it really? (The Presidential Address of the Modern Humanities Research Association.) *The Modern Language Review* 80(4): xxv–xxxv.
- Debreslioska, Sandra 2019. *Representing Discourse Referents in Speech and Gesture*. Lund: Lund University Press.
- Debreslioska, Sandra; Özyürek, Asli; Gullberg, Marianne; Perniss, Pamela 2013. Gestural viewpoint signals referent accessibility. *Discourse Processes* 50: 431–456.
- Efron, David 1972[1941]. *Gesture, Race and Culture*. (Ekman, Paul, preface.) The Hague: Mouton and Co.
- Ekman, Paul; Friesen Wallace V. 1969. The repertoire of nonverbal behavior: Categories, origins, usages and coding. *Semiotica* 1: 49–98.
- Fludernik, Monika 1996. *Towards a 'Natural' Narratology*. New York: Routledge.
- Forceville, Charles 2017. Visual and multimodal metaphor in advertising: Cultural perspectives. *Styles of Communication* 9(2): 26–41.
- Genette, Gérard 1980. *Narrative Discourse: An Essay in Method*. Ithaca. New York: Cornell University Press.
- Goldin-Meadow, Susan; Nusbaum, Howard; Kelly, Spencer D.; Wagner, Susan 2001. Explaining math: Gesturing lightens the load. *Psychological Science* 12(6): 516–522.
- Goodwin, Charles 2003. The body in action. In: Coupland, Justine; Gwyn, Richard (eds.), *Discourse, the Body and Identity*. Basingstoke: Palgrave Macmillan, 19–42.
- Green, Jennifer 2014. *Drawn from the Ground: Sound, Sign and Inscription in Central Australian Sand Stories*. Cambridge: Cambridge University Press.
- Gullberg, Marianne 2006. Handling discourse: Gestures, reference tracking, and communication strategies in early L2. *Language Learning* 56(1): 155–196.
- Haddington, Pentti; Kääntä, Leila (eds.) 2011. *Kieli, keho ja vuorovaikutus: Multimodaalinen näkökulma sosiaaliseen toimintaan*. (Suomalaisen Kirjallisuuden Seuran Toimituksia, Tiede, 1337.) Helsinki: Suomalaisen Kirjallisuuden Seura.
- Halliday, Michael A. K.; Hasan, Ruqaiya 1976. *Cohesion in English*. London: Longman.
- Hickmann, Maya; Hendriks, Henriëtte 1999. Cohesion and anaphora in children's narratives: A comparison of English, French, German, and Mandarin Chinese. *Child Language* 26: 419–452.

- Hockett, Charles F. 1960. The origin of speech. *Scientific American* 203: 88–111.
- Husserl, Edmund 1970[1936]. *The Crisis of European Sciences and Transcendental Phenomenology*. Evanston: Northwestern University Press.
- Hühn, Peter; Pier, John; Schmid, Wolf; Schönert, Jörg (eds.) 2009. *Handbook of Narratology*. Berlin: Walter de Gruyter.
- Jakobson, Roman 1959. Linguistic aspects of translation. In: Brower, Reuben A. (ed.), *On Translation*. Cambridge: Harvard University Press. Cambridge: Harvard University Press.
- 1965. Quest for the essence of language. *Diogenes* 13: 21–38.
- Jewitt, Carey 2014. Multimodal analysis. In: A. Georgakopoulou, Alexandra; Spilioti, Tereza (eds.), *The Routledge Handbook of Language and Digital Communication*. London: Routledge, 69–84.
- Jewitt, Carey; Bezemer, Jeff; O'Halloran, Kay 2016. *Introducing Multimodality*. London: Routledge.
- Jääskeläinen, Anni 2009. Mitä eletutkimuksessa tapahtuu? *Virittäjä* 4: 633–638.
- Kajannes, Katriina; Kirstinä, Leena 2000. *Kirjallisuus, kieli ja kognitio: Kognitiivisesta kirjallisuuden ja kielentutkimuksesta*. Helsinki: Helsinki University Press.
- Kendon, Adam 1980. Gesture and speech: Two aspects of the process of utterance. In: Key, Mary R. (ed.), *Nonverbal Communication and Language*. The Hague: Mouton, 207–227.
- 2004. *Gesture: Visible Action as Utterance*. Cambridge: Cambridge University Press.
- Kita, Sotaro 2009. Cross-cultural variation of speech-accompanying gesture: A review. *Language and Cognitive Processes* 24(2): 145–167.
- Kita, Sotaro; Gijn, Ingeborg van; Hulst, Harry van der 1998. Movement phases in signs and co-speech gestures, and their transcription by human coders. In: Wachsmuth, Ipke; Fröhlich, Martin (eds.), *Gesture and Sign Language in Human-Computer Interaction: International Gesture Workshop Bielefeld 1371*. Berlin: Springer, 23–35.
- Kita, Sotaro; Özyürek, Asli 2003. What does cross-linguistic variation in semantic coordination of speech and gesture reveal? Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and Language* 48: 16–32.
- Konderak, Piotr 2016. Between language and consciousness: Linguistic qualia, awareness, and cognitive models. *Studies in Logic, Grammar and Rhetoric* 48(61): 285–302.
- Kress, Gunther 2010. *Multimodality: A Social Semiotic Approach to Contemporary Communication*. London: Routledge.
- Labov, William 1972. *Language in the Inner City: Studies in the Black English Vernacular*. Oxford: Basil Blackwell.
- Labov, William; Waletzky, Joshua 1967. Narrative analysis: Oral versions of personal experience. In: Helm, June (ed.), *Essays on the Verbal and Visual Arts*. Seattle: University of Washington Press, 12–44.
- Lenninger, Sara 2016. Pictures: Perceptions of realism in the service of communication. In: Dunér, David; Sonesson, Göran (eds.), *Human Lifeworlds: The Cognitive Semiotics of Cultural Evolution*. Frankfurt am Main: Peter Lang Publishing Group, 97–112.
- Levy, Elena T.; McNeill, David 1992. Speech, gesture, and discourse. *Discourse Processes* 15: 277–301.
- Martinet, André 1984. Double articulation as a criterion of linguisticity. *Language Sciences* 6(1): 31–38.
- Mayer, Mercer 1969. *Frog. Where are You?* New York: Dial Press.

- McNeill, David 1992. *Hand and Mind: What Gestures Reveal about the Thought*. London: The University of Chicago Press.
- Merleau-Ponty, Maurice 2002. *Phenomenology of Perception*. (Smith, Colin, trans.) London: Routledge.
- Mushin, Ilana 2001. *Evidentiality and Epistemological Stance: Narrative Retelling*. Amsterdam: John Benjamin's Publishing Company.
- Prince, Gerald 2008. Narrativehood, narrativeness, narrativity, narratability. In: Pier, John; García Landa, José A. (eds.), *Theorizing Narrativity*. Berlin: De Gruyter, 19–27.
- Ranta, Michael 2011. Stories in pictures (and non-pictorial objects) – a narratological and cognitive psychological approach. *Contemporary Aesthetics* 9: 1–25.
- Rimé, Bernard; Schiaratura, Loris 1991. Gesture and speech. In: Feldman, Robert S.; Rimé, Bernard (eds.), *Fundamentals of Nonverbal Behaviour*. Cambridge: Cambridge University Press, 239–281.
- Sanford, Anthony J.; Emmott, Catherine 2012. *Mind, Brain and Narrative*. Cambridge: Cambridge University Press.
- Seppä, Anita 2012. *Kuvien tulkinta: Menetelmäopas kuvataiteen ja visuaalisen kulttuurin tulkitsijalle*. Tampere: Gaudeamus.
- Sonesson, Göran 1997. Mute narratives: New issues in the study of pictorial texts, Lund, May 1995. In: Lagerroth, Ulla-Britta; Lund, Hans; Hedling, Erik (eds.), *Interart Poetics: Acts of the Congress "Interart Studies: New Perspectives"*. Amsterdam: Rodopi, 243–252.
- 2010. From mimicry to mime by way of mimesis: Reflections on a general theory of iconicity. *Sign Systems Studies* 38(1/4): 18–66.
  - 2011. The mind in the picture and the picture in the mind: A phenomenological approach to cognitive semiotics. *Lexia: Rivista di semiotica* 07/08: 167–182.
  - 2014. Translation and other acts of meaning: In between cognitive semiotics and semiotics of culture. *Cognitive Semiotics* 7(2): 249–280.
  - 2015a. From remembering to memory by way of culture: A study in cognitive semiotics. *Southern Journal of Semiotics* 5(1): 25–52.
  - 2015b. Semiotics of photography: The state of the art. In: Trifonas, Peter (ed.), *International Handbook of Semiotics*. Dordrecht: Springer, 417–483.
- Stampoulidis, Georgios 2019. Stories of resistance in street art: A cognitive-semiotic approach. *Public Journal of Semiotics* 8(2): 417–483.
- Stampoulidis, Georgios; Bolognesi, Marianna; Zlatev, Jordan 2019. A cognitive semiotic exploration of metaphors in Greek street art. *Cognitive Semiotics* 12(1).
- Stöckl, Hartmut 2004. In between modes: Language and image in printed media. In: Ventola, Eija; Charles, Cassily; Kaltenbacher, Martin (eds.), *Perspectives on Multimodality*. Amsterdam: John Benjamins Publishing Company, 9–30.
- Todorov, Tzvetan 1969. Structural analysis of narrative. *NOVEL: A Forum on Fiction* 3(1): 70–76.
- Tomasello, Michael 2008. *The Origins of Human Communication*. Cambridge: MIT Press.
- Trask, Robert L. 2004. *Language: The Basics*. (2nd ed.) London, New York: Routledge.
- Vigliocco, Gabriella; Perniss, Pamela; Vinson, David 2014. Language as a multimodal phenomenon: Implications for language learning, processing and evolution. *Philosophical Transactions of the Royal Society B* 369(1651): 20130292.

- Wagner, Susan M.; Nusbaum, Howard; Goldin-Meadow, Susan 2004. Probing the mental representation of gesture: Is handwaving spatial? *Journal of Memory and Language* 50(4): 395–407.
- Wagner Cook, Susan 2014. Gesture and working memory. In: Müller, Cornelia; Cienki, Alan; Fricke, Ellen; Ladewig, Silva H.; McNeill, David; Bressemer, Jana (eds.), *Body – Language – Communication: An International Handbook on Multimodality in Human Interaction 2*. Berlin: De Gruyter Mouton, 1936–1940.
- Wittenburg, Peter; Brugman, Hennie; Russel, Albert; Klassmann, Alex; Sloetjes, Han 2006. ELAN: A professional framework for multimodality research. In: *Proceedings of LREC, Fifth International Conference on Language Resources and Evaluation, 1556–1559*.
- Yiheng Zhao, Henry 2015. Editor's note. *Signs & Media SCU* 11: 2–3.
- Zlatev, Jordan 2008. The dependence of language on consciousness. *Journal of Consciousness Studies* 15(6): 34–62.
- 2009. The semiotic hierarchy: Life, consciousness, signs and language. *Cognitive Semiotics* 4: 169–200.
  - 2012. Cognitive semiotics: An emerging field for the transdisciplinary study of meaning. *Public Journal of Semiotics* 4(1): 2–24.
  - 2015. The emergence of gestures. In: MacWhinney, Brian; O'Grady William (eds.), *The Handbook of Language Emergence*. Malden: Wiley-Blackwell Publishing, 458–477.
  - 2019. Mimesis theory, learning and polysemiotic communication. In: Peters, Michael (ed.), *Encyclopedia of Educational Philosophy and Theory*. Dordrecht: Springer.

**Appendix A.** List of Finnish clause initial connective devices found in the data and divided in five categories of conjunctive relation based on their function in the clause.

Additive	<p>eikä - 'nor'  eli - 'that is'  ja - 'and'</p> <p>ja kaiken lisäksi -  'and above all'  niin eli - 'so that is'</p> <p>tai - 'or'  tai sitten - 'or then'  vai - 'or'</p>		
Adversative	<p>että mutta - 'but that'  ja että - 'and that'  ja mutta - 'and but'</p> <p>että - 'that'  että kun - 'that when'  että sitten - 'that then'  ja että kun - 'and when'  ja koska - 'and because'</p>	<p>ja vaikka - 'and although'  mutta - 'but'  mutta että - 'but that'</p> <p>ja kun - 'and when'  ja niin - 'and so'  ja sitten koska -  'and then because'  ja sitten kun -  'and then when'</p>	<p>mutta sitten - 'but then'  vaan - 'but'</p> <p>ja sitten sen takia -  'and then because of that'  ja sitten siitä syystä -  'and then for that reason'  joten - 'so'  koska - 'because'</p> <p>koska sitten -  'because then'  kun - 'when'  mutta sitten kun -  'but then when'  niin että - 'so that'</p>
Causal	<p>eli ensin - 'so first'  ja aluksi - 'and at start'  ja edelleen - 'and still'  ja ensin - 'and first'  ja loppujen lopuksi -  'and in the end'  ja lopulta - 'and at last'  ja no sitten -  'and well then'  ja samaan aikaan -  'and at the same time'  ja samaan aikaan kun -  'and at the same time that'  ja samalla -  'and meanwhile'</p>	<p>ja samalla kun -  'and while'  ja sen jälkeen -  'and after that'  ja siinä samassa -  'and in that instance'  ja siinä vaiheessa -  'and at that point'  ja sillä aikaa -  'and meanwhile'  ja sillä välin sitten -  'and meanwhile then'  ja sitten - 'and then'  ja sitten samaan aikaan kun -  'and then while'</p>	<p>no sitten kun -  'well when'  sen jälkeen -  'after that'  siinä välissä -  'at that point'  sillä aikaa - 'meanwhile'  sillä välin - 'meantime'  sitten - 'then'  sitten kun -  'then when'</p>
Temporal	<p>että - 'that'  ja no - 'and well'  ja sitten no -  'and then well'</p>	<p>mutta kuitenkin -  'but anyway'  mutta kuitenkin että -  'but anyway that'</p>	<p>no - 'well'  no kuitenkin -  'well anyway'</p>
Continuative	<p>että - 'that'  ja no - 'and well'  ja sitten no -  'and then well'</p>	<p>mutta kuitenkin -  'but anyway'  mutta kuitenkin että -  'but anyway that'</p>	<p>no - 'well'  no kuitenkin -  'well anyway'</p>

**Appendix B. *Frog, Where Are You?* Transcript (English and Finnish)**

(Adapted from SALT Software LLC), 2015)

- Picture 1:** [There once was a boy who had a dog and a pet frog. He kept the frog in a large jar in his bedroom.] [Olipa kerran poika, jolla oli koira ja lemmikkisammakko. Hän piti sammakkoa isossa purkissa makuuhuoneessaan.]
- Picture 2:** [One night while he and his dog were sleeping, the frog climbed out of the jar. He jumped out of an open window.] [Eräänä yönä, kun poika ja koira nukkuivat, sammakko kiipesi pois purkista, ja hyppäsi ulos avoimesta ikkunasta.]
- Picture 3:** [When the boy and the dog woke up the next morning, they saw that the jar was empty.] [Seuraavana aamuna, kun poika ja koira heräsivät, he näkivät, että purkki oli tyhjä.]
- Picture 4:** [The boy looked everywhere for the frog. The dog looked for the frog too. When the dog tried to look in the jar, he got his head stuck.] [Poika ja koira etsivät sammakkoa kaikkialta. Kun koira yritti katsoa purkista, jäi sen pää purkkiin jumiin.]
- Picture 5:** [The boy called out the open window, “frog, where are you?” The dog leaned out the window with the jar still stuck on his head.] [Poika huusi avoimesta ikkunasta, ”sammakko, missä olet?” koiran nojatessa ikkunasta ulos purkki vieläkin päässään kiinni.]
- Picture 6:** [The jar was so heavy that the dog fell out of the window headfirst!] [Purkki oli niin painava, että koira tippui pää edellä ikkunasta ulos.]
- Picture 7:** [The boy picked up the dog to make sure he was ok. The dog wasn’t hurt but the jar was smashed.] [Poika nosti koiran syliinsä tarkistaakseen, että se oli kunnossa. Koiralle ei ollut käynyt kuinkaan, mutta purkki oli säpäleinä.]
- Picture 8:** [The boy and the dog looked outside for the frog. The boy called for the frog.] [Poika ja koira etsivät sammakkoa ulkona. Poika huusi taas sammakkoa.]
- Picture 9:** [He called down a hole in the ground while the dog barked at some bees in a beehive.] [Hän huusi maassa olevaan kuoppaan samalla, kun koira haukkui mehiläisille mehiläispesässä.]
- Picture 10:** [A gopher popped out of the hole and bit the boy right on his nose. Meanwhile, the dog was still bothering the bees, jumping up on the tree and barking at them.] [Taskurotta tuli esiin kuopasta, ja puri poikaa suoraan nenään. Samaan aikaan koira häiritsi vieläkin mehiläisiä, hyppi puuta vasten, ja haukkui niille.]
- Picture 11:** [The beehive fell down and all of the bees flew out. The bees were angry at the dog for ruining their home. The boy wasn’t paying any attention to the dog. He had noticed a large hole in a tree. So he climbed up the tree and called down the hole.] [Mehiläispesä tippui maahan, ja kaikki mehiläiset lensivät ulos pesästä. Mehiläiset olivat vihaisia koiralle kotinsa rikkomisesta. Poika ei kiinnittänyt mitään huomiota koiraan. Sen sijaan hän oli huomannut ison reiän puussa. Niinpä hän kiipesi puuhun, ja huusi reikään.]
- Picture 12:** [All of a sudden an owl swooped out of the hole and knocked the boy to the ground. The dog ran past the boy as fast as he could because the bees were chasing him.] [Yhtäkkiä pöllö syöksähti reiästä, ja tuuppasi pojan maahan. Koira juoksi pojan ohi niin kovaa kuin jaloistaan pääsi, koska mehiläiset jahtasivat sitä.]
- Picture 13:** [The owl chased the boy all the way to a large rock.] [Pöllö jahtasi poikaa aina suurelle kivelle saakka.]

**Picture 14:** [The boy climbed up on the rock and called again for his frog. He held onto some branches so he wouldn't fall.] [Poika kiipesi ylös kivelle, ja huusi taas sammakkoaan. Hän piti tukea oksista, jottei putoaisi.]

**Picture 15:** [But the branches weren't really branches! They were deer antlers. The deer picked up the boy on his head.] [Mutta oksat eivät oikeasti olleetkaan oksia! Ne olivatkin peuran sarvet. Peura nosti pojan sarvilleen,]

**Picture 16:** [The deer started running with the boy still on his head. The dog ran along too. They were getting close to a cliff.] [Ja lähti juoksemaan poika sarvissaan. Koira juoksi mukana myös. He lähestyivät jyrkännettä.]

**Picture 17:** [The deer stopped suddenly and the boy and the dog fell over the edge of the cliff.] [Peura pysähtyi äkkiä, ja poika ja koira tippuivat jyrkänteen laidan yli.]

**Picture 18:** [There was a pond below the cliff. They landed with a splash right on top of one another.] [Jyrkänteen alapuolella oli lammikko. Poika ja koira läiskähtivät toinen toisensa päälle lammikkoon.]

**Picture 19:** [They heard a familiar sound.] [Nyt he kuuluivat jostain tutun äänen.]

**Picture 20:** [They crept towards a big log and the boy told the dog to be very quiet.] [He hiipivät suuren kaatuneen puunrungon luo, ja poika käski koiraa olemaan hyvin hiljaa.]

**Picture 21:** [Together they peeked over the log.] [He nousivat yhdessä kurkistamaan puunrungon yli.]

**Picture 22:** [There they found the boy's pet frog. He had a mother frog with him.] [Sieltä he löysivät pojan lemmikkisammakon. Sillä oli toinen sammakko mukana.]

**Picture 23:** [They had some baby frogs and one of them jumped toward the boy.] [Niillä oli myös poikasia, ja yksi niistä hyppäsi poikaa kohti.]

**Picture 24:** [The baby frog liked the boy and the boy and the dog were happy to have a new pet frog to take home. As they walked away the boy waved and said "goodbye" to his old frog and his family.] [Vauvasammakko tykkäsi pojasta, ja poika ja koira olivat onnellisia saadessaan viedä uuden lemmikkisammakon kotiin. Kävellessään pois poika vilkutti, ja sanoi näkemiin vanhalle sammakolleen ja sen perheelle.]

**Appendix C. Feedback form**

(Translated from Finnish)

1. Gender:
2. Age:
3. Are you left- or right-handed?
4. Languages spoken rating from 'one' (beginner) to 'five' (advanced):
5. In your everyday life do you regularly use some other language in addition to Finnish (home environment/leisure/work/education)? If yes, what language and in which situations?
6. Have you lived abroad for longer periods than six months? If yes, where, how long, and at what age?
7. Did you consider the experiment challenging? If yes, why?
8. Do you think the task could have been easier or alternatively more challenging/seeing the story you would have heard the story spoken on tape? If yes, please indicate the reason for your answer.
9. Do you consider yourself a visual or an auditory learner (i.e. you learn better via visual perception, or auditory perception)?

Thank you for your participation!

## **Перевод с моносемиотического в полисемиотический нарратив: изучение финской речи и жестов**

Человеческая коммуникация может быть моносемиотической или полисемиотической, в зависимости от того, комбинируются ли репрезентаций из одной или более *семиотических систем*, таких как язык, жест и изображение. Каждая знаковая система имеет свои уникальные возможности рассказывания, что делает интерсемиотический перевод из одной системы в другую проблематичным. Мы исследовали влияние исходной семиотической системы, реализованной в речи и последовательности изображений на то, как та же самая история была пересказана с помощью речи и сопровождающими речь жестами. Рассказом служила история в картинках «Лягушка, где ты?» Одной группе показали историю в картинках, другая услышала ее в совпадающем устном рассказе. Каждый участник пересказывал свою историю и все рассказы записывались на видео и анализировались как с точки зрения речи, так и с точки зрения жестов. Учитывая высокую степень иконичности изображения, мы ожидали больше увидеть и иконических жестов в повествованиях, пересказанных с картинок, чем с устной речи. И наоборот, мы ожидали большей согласованности в нарративах, переведенных из речи. Результаты показали, что больше иконических жестов производилось в нарративах в переводе с речи, и, как и ожидалось, в повествованиях, переведенных из рассказа в картинках. Нарративы, произведенные участниками, которые только слышали историю, не имели большего разнообразия соединительных устройств, но тип устройств немного различался между группами. Результаты показывают, что история, представленная в различных семиотических системах, как правило, переводится в различные полисемиотические нарративы.

## **Monosemiotilise narratiivi tõlkimine polüsemiotiliseks: soome kõne ja žestide vaatlus**

Inimkommunikatsioon võib olla kas monosemiotiline või polüsemiotiline, olenevalt sellest, kas selles ühendatakse representatsioonivalikuid ühest või rohkematest semiootilistest süsteemidest nagu keel, žestid ja pildiline kujutamine. Igal semiootilisel süsteemil on omaenda unikaalne loovestmispotentsiaal, mis muudab intersemiootilise tõlke ühest süsteemist teise keerukaks. Uurisime kas kõnena või pildijadana realiseerunud semiootilise allüksüsteemi mõju sellele, kuidas sama lugu kõne ning kõnega kaasnevate žestide abil ümber jutustati. Looks oli pildiraamatu “Konn, kus sa oled?” sisu. Üks rühm soome keelt emakeelena kõnelejaid nägi lugu pildilises esituses ja teine kuulis sellega sobivat suulist jutustust. Kõik osalejad jutustasid loo adressandile ümber ning kõigest jutustustest tehti videosalvestused, mille kõnet ja žeste analüüsi. Võttes arvesse pildilise kirjeldamise kõrget ikoonilisuse astet, eeldasime piltidest tõlgitud narratiivide puhul rohkem ikoonilisi žeste (eriti läbimängu) võrreldes kõnest tõlgitud jutustustega. Vastupidiselt ootasime kõnest tõlgitud narratiivide puhul suuremat narratiivset koherentsust. Tulemustest nähtus, et kõnest tõlgitud narratiivides kasutati rohkem ikoonilisi žeste, ent enamasti ei kuulunud need läbimängimise alltüüpi. Nagu ootasime, oli ikooniline läbimäng sagedasem narratiivides, mida tõlgiti piltidena esitatud loost lähtuvalt. Narratiivides, mida esitasid lugu kuulnud osalejad, ei olnud konnektiivsed võtted mitmekesisemad, kuid rühmade vahel ilmnes võtete tüübis teatud erinevus. Arvestades ka mõningaid lisaerinevusi rühmade vahel, mida eelnevalt ei oodatud, näitavad tulemused, et erinevate semiootiliste süsteemidena esitatud lugu kaldutakse tõlkima erinevateks polüsemiotilisteks narratiivideks.