

A FRAME ANALYSIS OF AMERICAN SIGN LANGUAGE

JUDY ANNE KEGL AND NANCY CHINCHOR

*Department of Linguistics  
University of Massachusetts  
Amherst 01002*

ABSTRACT

This paper is a justification for the use of frame analysis as a linguistic theory of American Sign Language. We give examples to illustrate how frame analysis captures many of the important features of ASL.

0. Introduction

From a linguistic standpoint, we are interested in language processing systems for the claims that they make about language in general. Our interests in those claims leads us to examine what implications they may have for the analysis of languages other than English. The data from American Sign Language (ASL) is important because it is indicative of the way people perceive and represent events. This linguistic data requires careful analysis and much psychological insight before it can be used as evidence for any particular theory of representation of visual knowledge of events. We have tried to bring together some ideas from artificial intelligence, linguistics, and psycholinguistics in order to analyze the data from ASL.

The major framework we have adopted from AI is that of frames. Minsky's introduction of frames as a way of representing knowledge and the further

formulations of frames and related notions by Winograd and Fillmore form the bases for our frame analysis. We rely heavily on the work done by psycholinguists on visual perception as a justification for using frame analysis. Further justification comes as a result of the work of linguists and psycholinguists on ASL and the visual perception of the deaf.

The two most direct sources for our analysis of ASL are Reid (1974) and Thompson (1975). Reid's paper presents a clear and useful distinction between the linguistic level of the sentence and the conceptual level of the image. The sentence is a generalization and the image is an instantiation of that generalization. However, "the units in a sentence are not just realized as 'parts' of a whole represented in the image by the individual participants, rather these units act reciprocally to determine jointly the character of the related participants and to unite them into a system of dependencies." At the level of the sentence the verb is all-important because it governs the relations that exist between the nouns. However, it has no direct representation in the image; it is merely embodied in the structure of the image. Thompson's paper gives guidelines for using frames in linguistic analysis. His definitions of key concepts and his examples of frames for English have been a model for our analysis.

## 1. American Sign Language

ASL is the language of many deaf people in the US. There is a continuum encompassing the many version of several sign systems. ASL is a manual language composed of signs, fingerspelling, and occasional initialization of signs. It is in no way a signed version of English but is rather an independent language as different from English as is French or Japanese.

ASL is a visual language. This visual modality allows it not only a temporal but also a multidimensional spatial framework as well as freedom from many of the constraints normally put on a linear language. Many spatial relations can be preserved in miniature in what has been referred to in the sign literature as a visual analog. For example, the sentence, 'Fred stood in front of Harry,' does not necessitate a linear description. It can be represented by the indexicalized marker for FRED being positioned in the signing space in front of the one for HARRY. It is with respect to the specification of location and the use of deictic elements that sign most clearly distinguishes itself from spoken languages. This and other related problems in sign will be examined later in this paper. Focusing on the aspects of visual analog and deixis does not imply that sign does not employ many of the linear and temporal devices used in spoken languages, but rather that these devices serve different functions.

ASL is linearly ordered with respect to a standard method for presenting a scenario. The order of presentation is usually ground, then figures, then the action or relation involved. A room would be specified, then a door, then relevant furniture, then participants in an action. Generally, signs are presented in such a way as to allow further reference to them even if this **referencing** was not intended when the element was introduced into the discourse.

A relational grammar (Perlmutter and Postal) can be useful in describing ASL. Their grammar focuses on the relations of various participants in an action to the verb. The notion of subject can be related to what Friedman calls the Agent (AGENT-PATIENT) or what Reid calls the causer (CAUSER-AFFECTED ELEMENT-RANGE). The Agent or causer shows up in sign as the active participant,

the patient as the usually stationary participant being acted upon. As in relational grammar, these relations are based upon observational properties of the terms with respect to the verb. The relational model is attractive because it does not force one to specify the syntactic form of the sentence through a rigid ordering or tree structure.

Even more flexible is a frame analysis model which allows one to speak in terms of a scene or visual image. Proximal relations can then be preserved without translation into any linear forms. The frames approach emphasizes an important aspect so often repeated in descriptions of ASL. What one is doing is building a picture -- a scene. The signer is always thinking in terms of the picture he is presenting. He is trying to produce a miniature characterization of a real event. When elements of the event are present and within access for him to refer to in his discourse, he will use them. For example, he will point to an actual person rather than producing an arbitrary grammatical index to refer to that person. Describing sign language through frames allows one to stress the visual picture being presented. It allows also for the smooth integration of other communication conventions used within the speech act. For example, if mime is found to be more explicit than the use of conventionalized ASL forms, it can easily be incorporated into the discourse making the total presentation a more direct representation of the event.

## 2. Visual Logic

Boyes (1972) gives various arguments based on visual perception experiments for analyzing sign in terms of visual logic. By 'visual logic,' she means a system of rules similar to the rules people use to make sense of any

visual experience. In the next section we show that frame analysis can be considered an appropriate visual logic for sign language. First we would like to present the basic arguments from Boyes (1972) for using visual logic since these arguments also support the use of frame analysis.

There are three major results of visual perception experimentation which Boyes cites in order to begin a study of the constraints that the visual mode puts on a sign language. These results all show the limitations of visual memory as compared to auditory memory. These memory processes can each be divided into the same three stages. First, there is the initial storage of the stimulus which is identical to the actual stimulus. This part of memory is referred to as iconic memory (visual mode) or echoic memory (auditory mode). The next stage is short term memory where rehearsal can take place. Rehearsal is the process of repetition of the stored material during which the material is decoded, i.e., grouped into meaningful segments. This recoded material is then stored in long term memory.

One result that Boyes cites is that iconic memory is shorter than echoic memory. Iconic storage usually lasts for between 250 msec and 1 sec whereas echoic storage can last as long as 10 sec. A second fact is that the reaction time to visual stimuli is longer than that to auditory stimuli. The third result is that visual short term memory is more limited than auditory short term memory in that it does not seem to be able to hold as many items in the presence of continued input. The current figures for this are 4 or 5 items maximum in visual STM as opposed to  $7 \pm 2$  items in auditory STM. Boyes claims that this difference is due to the limited capacity for rehearsal of visual information.

All three of these results show that there is generally less time avail-

able for processing the sign sentence then there is for the spoken sentence. The temporal segmentation of sign would have to produce segments short enough to fit in iconic memory. And the sentence would have to be structured in such a way as to not tax STM with its limited rehearsal capacity. The sentence structure cannot rely on dependencies of elements which are temporally separated beyond the span of visual STM. Boyes seems to go a bit too far here and says that there should not be a "syntax which depends on decoding a temporal succession of images as a unit." But all this really means is that the sentences in ASL must be shorter than 5 items or that they must be processed in a way that does not require linguistic links between items which are separated by more than 4 items. Of course, more must be known about the linguistic processing of sign language before these conclusions can be made more specific.

In any case, it is clear that more information must be encoded per time interval in a visual language than in a spoken language, if we assume that the rate of transmission of information is to be the same in both. This can be accomplished by the mode of production in two ways. First, the symbol system used must be more direct, i.e., there should be a simpler mapping between visual sign and meaning than there is between sound and meaning. Secondly, sign must utilize its spatial dimensions to overcome the temporal limitations on the transmission of information. Frame analysis is able to represent these qualities of ASL.

### 3. Frame Analysis

Frames are a convention for representing knowledge. Frame analysis is a method for representing language as a system of frames. There are four

different types of linked frames that we will be using. These are discussed in Thompson (1975). Thompson attempts to resolve the apparent conflict in terminology with reference to the notions of scenes and frames in the work on prototype semantics (Fillmore and Rosch, MSSB, 1975) and the work on natural language understanding systems (Winograd and Bobrow, MSSB, 1975). In order to do so, he focuses in on two dichotomies. The first yields two types of frames, those representing knowledge of events and those representing linguistic knowledge. The second dichotomy further refines the categorization so that each type of frame can describe prototypic knowledge or knowledge of the instance at hand. These distinctions, then, give rise to four types of frames: Scene Prototype Frames (SPF), Scene Instance Frames (SIF), Linguistic Prototype Frames (LPF), and Linguistic Instance Frames (LIF). Before we discuss the structure of each type of frame we would like to indicate their possible functions in processing ASL. A sees an event and an SIF is formed with guidance from the appropriate SPF which was activated when one of its principle defining characteristics had been recognized. A wishes to communicate this scene to B. A constructs the sign sentences by following the links from the SPF to an LPF. The LPF will guide the filling in of an LIF based on the actual participants in the SIF thus producing the appropriate sign sentences. B watches A's signing and essentially reverses this process. An LIF begins to be formed and activates an LPF which guides the filling in of the LIF and causes the activation of an SPF. The SPF guides the filling in of the SIF with information from the LIF. Once the SIF contains all the requisite information, B is said to have understood what A signed to him.

What information do these frames contain and what are the various links, or "perspectives" as Thompson calls them, between these frames? Thompson

suggests a certain internal structure for these frames.

A frame contains at least three sorts of things: slots, states, and actions.

Slots are for identifying the participants in a given frame. Each slot has a name and a value. In an Instance Frame, these values will usually be names of other Instance Frames which describe the things which are filling each slot, while in Prototype Frames, they will usually be names of other Prototype Frames which contain information about the sort of thing which can fill the associated slot.

States are statements about various relationships which hold among the slots, and actions describe transitions between states.

We will need a slightly different structure because of the kind of information that is usually presented in sign. The major addition that we make is a category of slots called Ground which contains such things as the setting and the time element. We call the rest of the slots Figures. An example of an SPF would be {PREDATOR-PREY}.

{PREDATOR-PREY}

Slots

Ground

TIME {time}  
PLACE {place}

Figures

PRED {animal}  
PREY {animal}

States

- I. PRED doesn't have PREY
- II. PREY has protection
- III. PRED gets PREY
- IV. PREY gets caught

Actions

- A. I. becomes false and III. becomes true
  - B. II. becomes false and IV. becomes true
  - C. I. becomes true and IV. becomes false
  - D. II. becomes true and III. becomes false
- A or C, A implies B, C implies D

An instance of this frame would have the ground and figure slots filled in with links to other instance frames as in the following SIF..



{PREDATOR-PREY}

Slots

Ground

TIME {narrative time 413}

PLACE {house 584}

Figures

PRED {wolf 02}

PREY {pig 98}

States and Actions (as in SPF)

The corresponding LPF would look much the same except for the crucial addition of the verb. An LPF contains Ground and Figure slots along with a verb slot. The States and Actions are no longer present. Presumably the verb and the cases encode all this information. A perspective is given in order to match the Figure slots in the SPF with the case slots in the LPF.

{PREDATOR-PREY}

Slots

Ground

TIME {position on time line}

PLACE {position in sign space}

Figures

AGENT {'animal'}

PATIENT {'animal'}

VERB {'lex WANT,GET,EAT'}

Perspectives

{PREDATOR-PREY,SPF}

PRED = AGENT

PREY = PATIENT

This account of the LPF is much in the spirit of Thompson's LPF. But our account of the LIF is different. We are dealing with sign and not a spoken language. The case relations are clearly manifested on the surface in sign because the hands act out the scene. So our LIF looks as follows:

{PREDATOR-PREY}

Slots

Ground

TIME {position on time line 617}

PLACE {position in sign space 729}

Figures

AGENT {wolf 44}

PATIENT {pig 91}

VERB {WANT,GET,EAT}

There is no need to have Thompson's perspective to tell us what case roles

the subject, object, etc. of the verb play in the prototype. Processing will be faster since the linguistic prototype and instance frames are more alike in ASL.

In sign the four frames are more alike in structure and there is much less need for links between frames. This cuts down processing time greatly and compensates for the limitations on visual memory. Linguistic frames differ from scene frames in the presence of the verb. As Reid says, the grammar of the image is different from the grammar of the language in that the image is made up of participants and properties attributed to them whereas the sentence is a package held together by the verb. Frame analysis formalizes this notion and reflects the speed of processing ASL. We propose that it be seriously explored as a linguistic theory for sign language.

#### 4. A Frame Analysis of Sign Language

The remainder of this paper will include a description of some devices in sign as well as a discussion of how they might be handled by a theory of Frame Analysis. These devices are not only interesting features to analyze, but also reveal the structure of the frames (focus, boundaries, weak points).

Indexing is a process in ASL which parallels pronominalization and deixis (this, that, here, there) in spoken language. There are two types of indexing: real world references and conventional references.

Real world references are of the type discussed earlier. When the person referred to is in the vicinity, one points directly to that person rather than to an arbitrary index. The same goes for location. Also, a person recently having left a group of signers will be referred to by pointing to the position he previously occupied.

In frame analysis, the grammatical to real world reference link could be achieved by resorting to a higher frame encompassing the speech act. This speech act frame monitors the entire event and specifies what is common knowledge shared among the participants in the speech act. That shared knowledge determines the set of objects, persons and locations which can be referred to directly (by means of pointing). For example, if A knows that B has in his knowledge of the room they are in the vision of a bookshelf in one corner, then A can point directly to it without having to name it. The same goes for the shared knowledge of locations. If two people share the knowledge that city X is the obvious referent of a point back over the left shoulder, then it will be used. Where this knowledge isn't shared, this referencing would be forbidden.

There are several types of conventional indices for things, locations and people as well as positions for such indexing. The stationary person index, commonly referred to as grammatical indexing, involves referring to certain individuals by pointing to conventional places within the signing space: right, left, distal right, distal left, and straight ahead, in that order (for a right-handed signer). Indexing into these positions allows ready reference at any following time within the discourse.

Grammatical indexing uses a frame for reference similar to the speech act frame. In this frame, however, index points are specified as to which arbitrary referents are tied to them. In cases where participants are closely linked to spatial locations, they use these locations as their index points.

Indices must be established (i.e. JOHN (indexed left position); ALICE

(indexed right position)). Since the tie between these indices and their referents is weak and arbitrary, they must frequently be reestablished. In the videotape, reindexing played a role in aiding us in our determination of frame boundaries. Reindexing interacts with the sign we have termed NEUTRAL POSITION (arms drop to sides). NEUTRAL POSITION is used to mark the end of a long discourse. Directly following NEUTRAL POSITION, at the beginning of a new frame, the signer would reindex 3 (the sign THREE) and focus upon one of the three pigs. Reindexing also marks mistakes and overcomplicated referencing.

Besides NEUTRAL POSITION, there is another PAUSE SIGN which aids in the delineation of discourse and, therefore, in the discovery of frames. The PAUSE SIGN occurs at breaks between actions within frames or at shifts between agentive characters in frames.

Other key sign structures which aid in frame determination are body position shifting and the use of index markers. As a result of the limited length of this paper we cannot fully examine these devices here. However, an extended version of this paper and copies of the transcription of the videotape of "The Three Little Pigs" are available from the authors.

### Acknowledgements.

We would like to thank Tommy Radford for his help both in the signing of the story of The Three Little Pigs and in providing helpful comments for its analysis. We would also like to thank the Sign Group and the Frames Group from the MSSB summer meetings, Berkeley, 1975. A special note of thanks to George Lakoff whose insights into our common interests made this paper possible. The research for this paper was supported by the 1975 MSSB Workshop on Alternative Theories of Syntax and Semantics.

### Bibliography.

- Boyes, Penny. 1972. "Visual Processing and the Structure of Sign Language." unpublished ms.
- Friedman, Lynn. 1975. "On the Semantics of Space, Time, and Person Reference in the American Sign Language." unpublished Master's Thesis, University of California at Berkeley.
- Reid, L. Starling. 1974. "Toward a Grammar of the Image." *Psychological Bulletin*, vol. 81, no. 6 (June), pp. 319-334.
- Thompson, Henry. 1975. "Frames for Linguists." unpublished ms.

END

