Analysis of ASL Motion Capture Data towards Identification of Verb Type

Evguenia Malaia
John Borneman
Ronnie B. Wilbur
Purdue University (USA)
email: emalaya@purdue.edu

Abstract

This paper provides a preliminary analysis of American Sign Language predicate motion signatures, obtained using a motion capture system, toward identification of a predicate's event structure as telic or atelic. The pilot data demonstrates that production differences between signed predicates can be used to model the probabilities of a predicate belonging to telic or atelic classes based on their motion signature in 3D, using either maximal velocity achieved within the sign, or maximal velocity and minimal acceleration data from each predicate. The solution to the problem of computationally identifying predicate types in ASL video data could significantly simplify the task of identifying verbal complements, arguments and modifiers, which compose the rest of the sentence, and ultimately contribute to solving the problem of automatic ASL recognition.

1 Introduction

In recent work, we have provided preliminary data indicating that there is a significant difference in the motion signatures of lexical predicate signs that denote telic and atelic events (Wilbur and Malaia, 2008b,c,a; Malaia et al., 2008). These results are empirical evidence for direct mapping between sign language (ASL) phonology/kinematics, and semantic decomposition of predicates (the Event Visibility Hypothesis, or EVH (Wilbur, 2003, 2009)). The present paper reviews this analysis of ASL predicate 3D motion signatures and considers further application of such data for computational processing of ASL video streams, and automatic recognition of predicate type based on 2D motion signatures. Particular attention is paid to the contribution of the slope of deceleration at the end of signs, and to the values of the maximum velocity and minimum acceleration achieved during the sign motion. The focus of the study on predicates is determined by the fact that each sentence or clause in natural languages is built around a predicate. Thus, a solution to the problem of identifying predicate types could significantly simplify the task of identifying verbal complements, arguments and modifiers, which compose the rest of the sentence.

2 Modeling events in ASL

Linguistic theory of verbal types has long observed universal correspondences between verbal meaning and syntactic behavior, including adverbial modification (Tenny, 2000), aspectual coercion (Smith, 1991), and argument structure alternations (Levin, 1993; Ramchand, 2008). Vendler (1967) proposed a system of four basic syntactically relevant semantic types of predicates: atelic States and Activities, and telic Achievements and Accomplishments. The telic/atelic distinction is most clearly analyzed in terms of the internal structure of events. 'Telic' is understood as the property of linguistic predicates (events) containing a conceptual (semantic) endpoint. In contrast, 'atelic' events do not contain such a point and have the potential to continue indefinitely. Atelic events are homogenous, in that they may be divided into identical intervals, each of which is an instance of the event itself, i.e. 'walking' as an instance of 'walking'. Telic events are composed of at least two sub-events, one of which is the final state, and are therefore heterogeneous (cannot be divided into identical intervals). The model was further developed by Pustejovsky (1991), with the primary distinction between static sub-event type S(tate) and dynamic sub-event type P(rocess). Telic events with transitions to the final state were modeled as combinations of non-identical sub-events (Table 1).

Table 1: Pustejovsky's predicate typology

Predicate type	**Definition**
Activity	P
State	S
Accomplishment	P $\rightarrow$ S
Achievement	S $\rightarrow$ S

Most recently, Ramchand (2008) has taken an event as the basis for hierarchical composition of phrases that replace the traditional notion of Verb Phrase, thereby

simplifying the interaction between the lexicon and the syntax, at least for dynamic events. This simplification has the potential to be very useful for automatic recognition of predicate signs. Ramchand divides events into a maximum of three hierarchical phrases: an initiation phrase (InitP), a process phrase (ProcP), and a result phrase (ResP). Each of these has an associated participant: InitP: Initiator; ProcP: Undergoer; and ResP: Resultee. This eliminates traditional problems associated with determining argument structure and thematic role assignment. One or more of these phrases may be identified by a single morpheme/word/sign. As a result, the same event could be represented by one word or an entire phrase, depending on the morphology of a particular language. Ramchand further demonstrate that expression of degrees of causation (direct or indirect) is related to whether a single morpheme identifies both [init] and [proc] (yielding interpretation as direct causation) or separate morphemes are needed (yielding indirect causation). Similar effects with resultatives are found with single morpheme identification of [proc] and [res] as compared to separate morphemes.

From this perspective we can analyze ASL signs in terms of the phrases they identify. In this paper we compare signs which identify at least [ResP] (telic events) with those that do not (atelic events).[1] The notion of event-based analysis of sentential semantics and syntax was supported in general for ASL in Rathmann (2005), and for Austrian Sign Language (Schalber, 2004, 2006). Semantics of event type has been shown to have a direct effect on available morphological modifications as well: Brentari (1998) notes that [delayed completive] aspect marking only applies to telic stems. Wilbur (2009) demonstrates that some types of aspectual marking (continuative, durative) can only apply to atelic predicates. Wilbur (2003) argued that the phonological structure of predicate signs in ASL shows event composition, and that the components are grammaticalized from universally available physics of motion and geometry of space. This Event Visibility Hypothesis (EVH) was formalized as 'movement which stops at points (p) in space maps semantically to the final State of telic events (en) and its individual argument semantic variable (x)'. In Ramchand's terms, ResP can be seen in lexical predicates representing telic events by the way the movement comes to a stop. This hypothesis was tested in the motion capture experiment described below.

3 Materials and methods

A group of 29 telic and 21 atelic ASL signs were randomized, and presented as a list via Powerpoint five times through. A native bilingual right-handed ASL signer wore a Gypsy 3.0 wired motion capture suit (Figure 2).

The signer viewed the powerpoint slides with stimuli and produced the list twice with each sign in isolation, once with each sign in the carrier phrase 'SIGN X AGAIN', and once sentence- medially 'SHE X TODAY'. For each production the hands began at rest, were raised for signing, and were returned to rest. We report the data from the marker on the right wrist, as the dominant right hand carries most of the meaningful motion information in ASL (the non- dominant typically serves as ground or repeats the movement of the dominant hand). All signs selected for the experiment

[1] As Ramchand notes, it is possible to get telic readings without ResP from bounded path complements. However this study uses lexical items, for which bounded path analyses have not yet been demonstrated, thus we treat them all as ResP items for expository purposes.

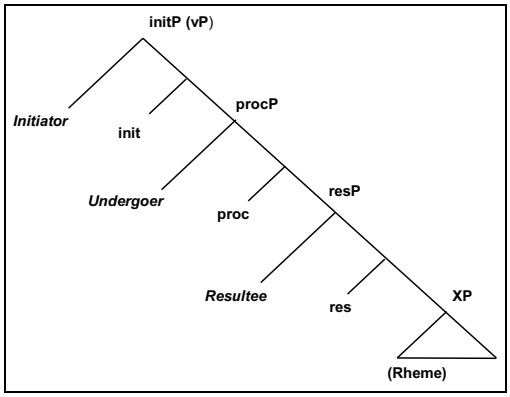

Figure 1: Ramchand's Event Structure projection

Figure 2: Signer in motion capture suit

included motion of the wrist. The data from the motion capture suit was recorded into Motionbuilder software and exported as text data for further analysis. The data included frame numbers, and marker positions along the 3D axis in millimeters for all recorded frames. Acquisition rates were 50 fps for kinematic data, and 30 fps for video data; the video also included an audio marker for alignment of the beginning of motion capture recording with the separate video recording. The time course of predicate signing in the video was annotated using ELAN software (Max Planck Institute for Psycholinguistics). The beginning of each predicate was marked once the dominant hand assumed appropriate handshape, and the end of each movement was marked at the final point of contact or maximal displacement in the lexical form of the sign. All start/stop times were determined by a single coder with over two decades experience measuring sign movement, with +/– 1 video frame precision. The vectors for the 3D location of the wrist marker were then imported into ELAN and aligned with video data using the audio marker. In addition to raw displacement data, derivatives of speed (in m/s) and acceleration (in m/s^2) were calculated in MATLAB and imported into ELAN. To minimize the difference in acquisition rates for video and kinematic data, velocity and acceleration vectors were imported into ELAN, and peak changes corresponding to the actual motion capture data points were compared to the annotations. This alignment was used to ensure the proper extraction of the motion capture data corresponding to the target sign between the marked start and end locations. Additionally, the error of measurement for each predicate was considered, evaluated as ratio of frame duration vs. predicate duration expressed in percentage. Consequently, predicates which spanned fewer than three video frames were discarded because of high error margin. For the rest of the predicates, the following metrics were calculated:

- the maximal velocity (maxV);

- the local minimum velocity following maxV (minV);
- the slope of the drop from maxV to minV (slope);
- the minimum acceleration (minA) following the maximum velocity;
- duration of the predicate (in frames);
- the frame location of maxV, minV, and minA.

These metrics were chosen to allow for maximal homogeneity in predicate comparison. The metrics related to the start of the sign were avoided as linguistically unreliable, while using the local velocity minima mitigated the effect of data interpolation in ELAN resulting from the frame difference between video recording and motion capture recording. The data were submitted to SPSS multivariate ANOVA to determine the effect of telicity value (Telic vs. Atelic).

4 Linguistic results and observations

The data from the right wrist marker indicate that in all environments, the deceleration of telic signs is steeper than that of atelic signs (Table 2).

Table 2: Deceleration slope for telic and atelic signs (* $p < 0.05$; ** $p < 0.001$)

Deceleration (mm/s2)	**Atelic mean (tokens used)**	**Telic mean (tokens used)**	**Telic/Atelic Ratio and effect size**
Isolation 1	–0.093 (13)	–0.136 (22)	1.46* F(1)= 4.528
Isolation 2	–0.123 (17)	–0.179 (23)	1.46* F(1)= 5.709
Carrier Phrase	–0.118 (14)	–0.233 (22)	1.97** F(1)= 15.258
Sentence	–0.14 (13)	–0.23 (18)	1.62* F(1)= 7.400

The data support the Event Visibility Hypothesis in ASL, indicating that there is a production difference reflecting the semantic distinction of event type in predicates. It appears that ASL takes advantage of available perceptual distinctions to provide cues to the viewer regarding the semantics of the predicate. Telic predicates in general have a steeper deceleration slope, marking the end-state of telic events. This deceleration may correspond to what Klima and Bellugi (1979) referred to as 'end marking'. From the perspective of syntax-semantics interface modeling theory (Ramchand, 2008; Wilbur, 2003), higher decelerations in motion signatures of telic ASL predicates also mark additional semantic arguments of the event, what Ramchand refers to as the 'Resultee'.

5 Development of metrics for computational identification of predicate types in ASL

The data from the four productions was compared in order to evaluate the consistency of production. The interval distribution of maxV and minA for all predicates, in both carrier phrases and in sentences, overlap (Figure 3), indicating that those two production conditions were not significantly different, and therefore this data could be pooled for the purposes of statistical analysis.

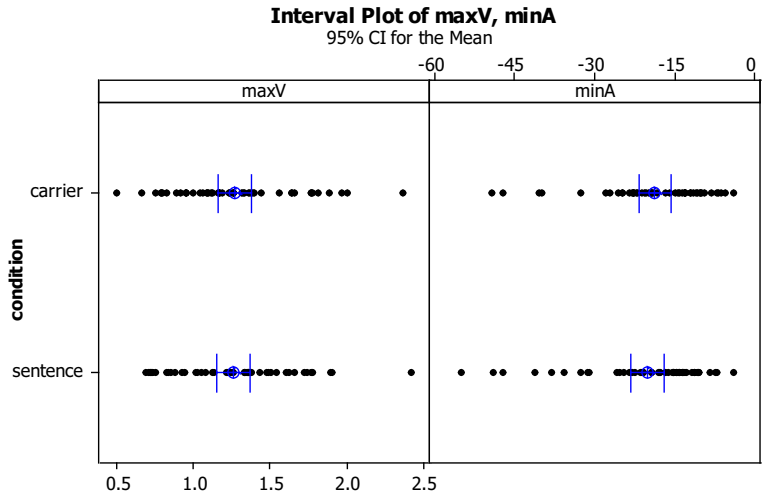

Figure 3: Interval plots of maxV and minA in carrier phrases and sentences for all predicates, displaying 95% confidence intervals for the mean values of the respective metrics

A similar comparison of maxV intervals by predicate type of those produced in isolation (Figure 4) revealed significant discrepancies between the two instances of production for atelic signs, possibly related to the attempt by the signer to reproduce the vocabulary form of the predicate. Thus, isolation data was not used for the following analysis.

A binary logistic regression was performed on the pooled data from both carrier phrase and sentence sign production, in order to search for the minimum number of variables that could be used to predict whether the signed predicate is telic or atelic. For this analysis the logit function was selected to calculate the predicted probabilities. Regression was first run with all of the measured variables from Section 3 included in the model. The variables with p-values above 0.05 threshold were rejected one at a time, and the regression calculation re-run; the process was repeated until all predictive variables in the model were below p=0.05. The model was reduced to two significant predictors: maxV and minA (maximum velocity and deceleration). The final regression model is shown in equation (1) using the variable dependence shown in equation (2).

$$P_T = \frac{e^{\beta}}{1+e^{\beta}} \tag{1}$$

$$\beta = -4.46 + 2.63\,(\text{maxV}) - 0.097\,(\text{minA}) \tag{2}$$

P_T is the probability of a predicate being telic, based on measured values for maxV and minA. Applying the above equation on the pooled data for carrier phrase and sentence production conditions (setting a 50% threshold so that $P_T > 0.5$ predicts telic, and $P_T < 0.5$ predicts atelic) ensures that 46 out of 56 telic predicates, and 32 out of

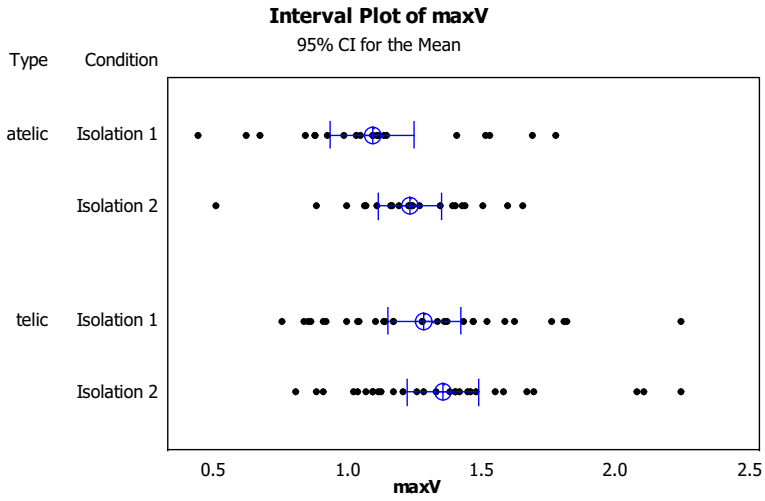

Figure 4: Interval plots of maxV in isolation productions 1 and 2 by predicate type

44 atelic predicates can be identified correctly using only maxV and minA measures (Figure 5).

However, comparing the coefficients of maxV (2.63) and minA (–0.097) in equation (2) makes it apparent that the effect of the deceleration (minA) on the output of the model is low. A regression analysis using only the maxV measurement to determine the predicate type yields a revised beta value as shown in equation (3).

$$\beta = -4.19 + 3.71\ (\text{maxV}) \tag{3}$$

Using this simplified equation on the pooled data ensures that 47 out of 56 telic predicates, and 27 out of 44 atelic predicates can be identified correctly with a 50% probability threshold based only on maxV (Figure 6).

Direct analysis of the maximum velocity (maxV) data supports both the original model (eqn. 3) and the simplified model (eqn. 4). Telic predicates have a significantly higher maxV mean and distribution than atelic predicates, as shown in Figure 7. It is this difference in the velocity distribution that allows for telic/atelic predictions based on maxV measurements.

6 Conclusion

The above pilot data analysis indicates that there exists a production difference in maximal velocity and deceleration slope of ASL predicate signs reflecting semantic distinction of event type in ASL predicates. From the linguistic standpoint, the overt difference in sign production maps onto an event-structural representation for the syntax-semantics interface, which has implications for modeling the syntax-semantics interface in both signed and spoken languages. Empirical evidence for Event Visibility in signed languages demonstrated that individual meaningful features in signs (such as rapid deceleration to a stop) can combine to create patterns which merge the syntactic

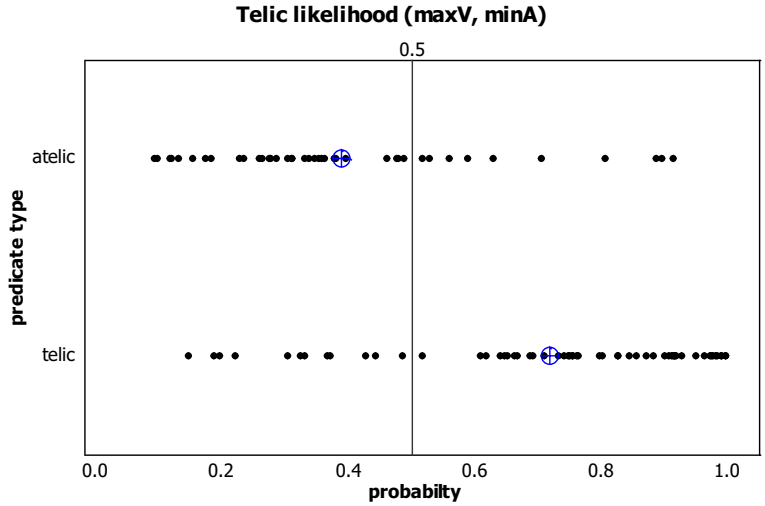

Figure 5: Probabilities of correct predicate type identification based on maxV and minA data. Data points represent the telic and atelic predicates, presented according to the probablility of their correct identification using equation (1) and variable dependence in equation (2)

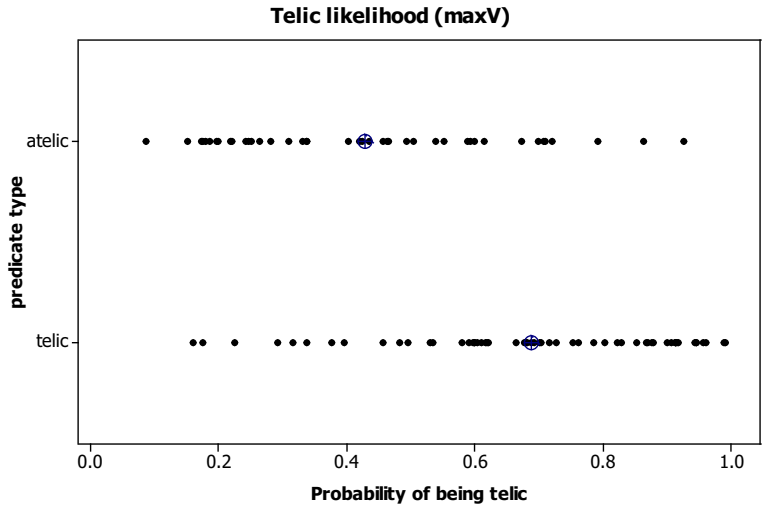

Figure 6: Probabilities of correct predicate type identification based on only maxV data. Data points represent telic and atelic predicates, presented according to the probability of their correct identification using revised beta value in equation (3). The crosshairs represent the mean probability of correct predicate type identification for atelic (0.429) and telic (0.689) predicates

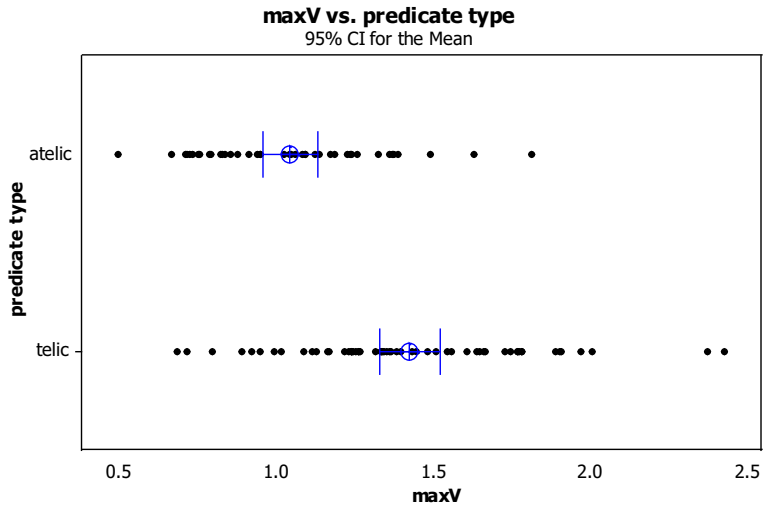

Figure 7: MaxV distribution for telic and atelic predicates in carrier phrases and sentences (pooled), displaying the mean values and 95% confidence intervals

level of a sign with its phonological level — the phenomenon which can be utilized for machine translation of signed languages.

For the purposes of computational approaches to sign recognition, the pilot data demonstrates that production differences between predicate types can be used to model the probabilities of specific predicate types occurring within the motion signature, based on either maximal velocity achieved within the sign, or maximal velocity and minimal acceleration data from each predicate. However, as pilot data analysis shows, higher acquisition rates for video and motion capture data would be beneficial to take full advantage of production differences in velocity and deceleration of different types of ASL predicates. Further research is needed to determine inter-signer variability in production differences between telic and atelic predicate signs, the reliability of maximal velocity and minimal acceleration metrics, and development of additional metrics (possibly similar to ones used for spoken language phonology (Adams et al., 1993)) which could rely on higher temporal resolution in data acquisition.

Acknowledgments Motion capture was conducted at the Envision Center for Data Perceptualization at Purdue University. We are grateful to Robin Shay, Gabriel Masters, Nicoletta Adamo-Villani, and the Purdue and Indianapolis sign language communities for their ongoing support of the Purdue Sign Language Linguistics Lab research. This work was supported by NSF Research in Disabilities Education grant #0622900 and by NIH grant DC00524 to R.B. Wilbur.

References

Adams, S., G. Weismer, and R. Kent (1993). Speaking rate and speech movement velocity profiles. *Journal of Speech and Hearing Research 36*, 41–54.

Brentari, D. (1998). *A prosodic model of sign language phonology*. Cambridge, MA: MIT Press.

Klima, E. and U. Bellugi (1979). *The Signs of Language*. Cambridge, MA: Harvard University Press.

Levin, B. (1993). *English Verb classes and alternations*. The Univ. of Chicago Press.

Malaia, E., R. Wilbur, and T. Talavage (2008). Experimental evidence of event structure effects on asl predicate production and neural processing. In *Proceedings of the 44th meeting of Chicago Linguistic Society*, Chicago, IL.

Pustejovsky, J. (1991). The syntax of event structure. *Cognition 41(1-3)*, 47–81.

Ramchand, G. (2008). *Verb Meaning and the Lexicon: A First Phase Syntax*. Cambridge: Cambridge University Press.

Rathmann, C. (2005). *Event Structure in American Sign Language*. Ph. D. thesis, University of Texas at Austin.

Schalber, K. (2004). Phonological visibility of event structure in Austrian Sign Language: A comparison of ASL and ÖGS. Master's thesis, Purdue University.

Schalber, K. (2006). Event visibility in Austrian Sign Language (ÖGS). *Sign Language & Linguistics 9*, 207–231.

Smith, C. (1991). *The Parameter of Aspect*. Dordrecht: Kluwer Academic Publishers.

Tenny, C. (2000). Core events and adverbial modification. In Tenny and Pustejovsky (Eds.), *Events as grammatical objects*. Stanford, CA: CSLA Publications.

Vendler, Z. (1967). *Linguistics in Philosophy*. Cornell University Press, New York.

Wilbur, R. (2003). Representations of telicity in ASL. *Chicago Linguistic Society 39(1)*, 354–368.

Wilbur, R. (2009). Productive reduplication in ASL, a fundamentally monosyllabic language. to appear in Language Sciences.

Wilbur, R. and E. Malaia (2008a). Contributions of sign language research to gesture understanding: What can multimodal computational systems learn from sign language research. *International Journal of Semantic Computing 2(1)*, 1–15.

Wilbur, R. and E. Malaia (2008b). Event Visibility Hypothesis: motion capture evidence for overt marking of telicity in ASL. In *Linguistic Society of America Annual Meeting*. Chicago: LSA.

Wilbur, R. and E. Malaia (2008c). From Encyclopedic Semantics to Grammatical Aspects: Converging Evidence from ASL and Co-Speech Gestures. In *DGfS annual meeting (AG 11, Gestures: A comparison of signed and spoken languages)*. Bamberg, Germany: DGfS.